

#### CANDIDATE GENE ANALYSIS OF THE 7DS QTL FOR DWARF BUNT RESISTANCE USING TARGETED CAPTURE SEQUENCING

Guriqbal Dhillon, Pabitra Joshi, Yulin Yan, Fangming Xiao, Will Krause, Magaret Krause, and **Jianli Chen** 

# Ι

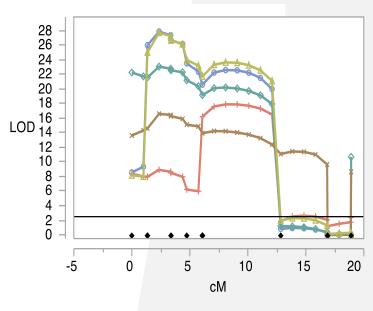
# **IMPORTANCE OF BUNT DISEASES**



- Dwarf bunt caused by *T. controversa and* common bunt caused by *T. laevis* Kühn and *T. tritici* (Bjerk.) Wint *are two destructive diseases in wheat*
- Cause significant yield losses and quality reduction
- Assessment of the disease is difficult and expensive
- Developing resistant wheat varieties crucial for sustainable crop management
- Requires a thorough understanding of the underlying genetic mechanisms
- Genomic analysis has emerged as a powerful tool to identify and characterize genes involved in resistance to bunt diseases

# Ι

#### A MAJOR QTL QDB.UI-7D FOUND BY CHEN ET AL., 2016



- Mapping population developed using IDO444 and Rio Blanco
- 159 RILs phenotyped across 4 environments in Logan, UT
- Genotyped using DArT markers and 9K SNPChip
- Three QTLs identified on chromosomes 1A, 2B, and 7DS
- QTL QDB.ui-7D detected with LOD 18.2 and R<sup>2</sup> 43.4%

• Another QTL *QDB.ufa-7D* was also reported just downstream of the *QDB.ui-7D* 5 Mb apart (Muellner et al., 2020)

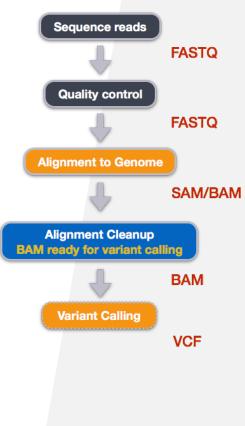


### **RESEARCH OBJECTIVES**

- To use capture sequencing in a targeted region flanking the major QTL QDB.ui-7D
- To use bioinformatic tools to analyze and identify candidate genes underlining the QDB.ui-7D
- To validate the candidate genes using loss of function and gain of function experiments
- To develop the resistant cultivars using genomics assisted technologies

# Ι

## **TARGETED CAPTURE SEQUENCING**



- A 3-11 Mb region of the 7DS was targeted using Illumina PE151 paired-end sequencing.
- Resistant parent 'IDO444' and susceptible parent 'RioBlanco' plus 29 other lines were sequenced.
- > More than 10 million reads per sample obtained.
  - Low quality bases and adapter sequences removed.
  - ➢ Filtered sequence mapped on refseqV1.0 using the BWA.
- > Resulting raw BAM files were filtered to region of interest.
- ➤ Resulted in an average of 2.72 million reads per sample.



# **READ MAPPING STATISTICS**

Sample	IDO444	IDO444 filtered	RioBlanco	RioBlanco filtered
reads mapped	18,140,184	2,736,437	16,352,209	2,284,793
reads mapped percent	99.820	99.784	99.787	99.744
raw total sequences	18,172,890	2,742,362	16,387,086	2,290,663
error rate	0.011	0.016	0.012	0.017
non primary alignments	0	0	0	0
reads MQ0 percent	7.824	7.391	8.275	7.118
reads properly paired percent	95.665	96.344	95.381	96.079

# 57

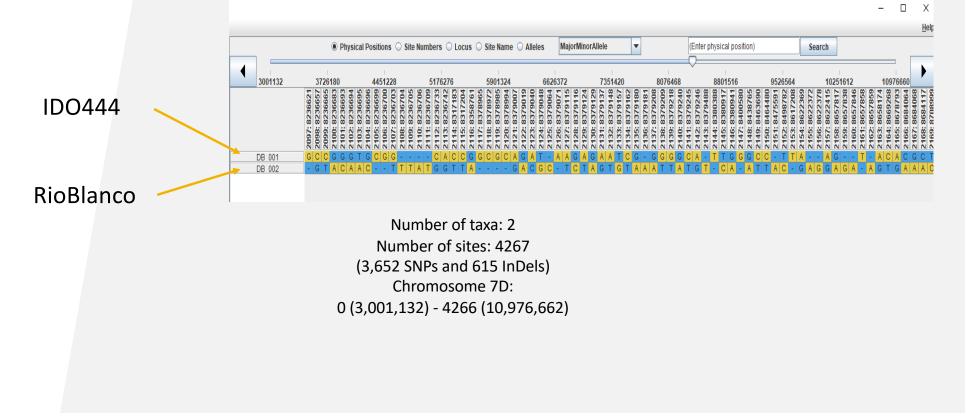
### **RAW VARIANTS:** 94647 SITES FROM 3,000,149 - 10,986,839

			Physi	cal Positions 🔘 Si	te Numbers 🔘 Locu	ıs 🔾 Site Name 🔾 Alle	eles MajorMinorA	Allele		Enter physical position)	S	earch	
													[
		3000149	3726211	4452273	5178335	5904397	6630459	7356521	8082583	8808645	9534707	10260769	10986831
			2 8827824 2 8827828 2 8827829 2 8827829 2 8827844 2 8827844 2 8827844 2 8827844 2 8827844	: 8827850 : 8827865 : 8827862 : 8827869 : 8827869 : 8827889 : 8827889 : 8827889	8827896 8827897 8827957 8827957 8827957 8827957 8827957	8827995 8828011 8828014 8828014 8828019 8828020 8828021 8828032 8828035	8828057 8828061 8828068 8828068 8828074 8828082 8828082 8828082 8828086 8828086	8828143 8828146 8828153 8828154 8828155 8828155 8828155	184 188 197 201	09 115 117 117 21 21 21 47	8828269 8828269 8828270 8828275 8828283 8828283 8828283	88283291 8828328 8828331 8828335 8828335 8828335 8828339 8828339	8828340 8828355 8828361 8828435 8828440 8828440 8828455
	Genotype		70268 70269 70271 70271 70272	70274 70275 70276 70276 70278 70279	70281 70281 70283 70283 70284 70284 70285	70287 70289 70289 70290 70291 70291 70293 70293	70295 70296 70297 70298 70299 70299 70301	70302 70303 70304 70305 70305	70308 70309 70310 70310 70312	70313 70314 70315 70315 70316 70318 70318 70318 70318	70321 70322 70323 70324 70325 70325 70326	70327 70328 70329 70330 70331 70332	70333: 70334: 70335: 70336: 70336: 70338: 70338:
	Genotype	DB 001	TTTGA-	KCTTGG	TGGGAG-	TCTRATCC	Y - C T A C T	CGCAGG	GGCCC	GTAT-NGC	CGGCR-	СТСТСС	CCTCCA
		DB 002	WKYRR -	TWKYOR	YSRGAG-	TTCGGCCCC	C - C T A C T	CACTGT	GATGT	G - T C T G G T	CTTGGC	- TACCT	CTCTAG
	IDOAAA	DB 003	WKYRR -	TWKYOR	YSRGAG-	TTCGGCCC	C - C T A C T	CACTGT	GATGT	G - T C T G G T	CTTGGC	- TACCT	CTCTAG
	IDO444	DB 004	TTTGA-	KCTTGG	TGGGAG-	TCTRATCC	Y - C T A C T	CGCAGG	GGCCC	GTAT-NGC	CGGCR-	CTCTCC	CCTCCA
		DB 005	T T T G A -	K C I I G G	I G G G A G -		Y - C T A C T	CGCAGG	GGCCC	GTAT-NGC	CGGCR-	CICICC	CCTCCA
		DB 006	TTTGA-	K C I I G G	I G G G A G -		Y - C T A C T	CGCAGG	GGCCC	GIAI-NGC	CGGCR-	CICICC	CCTCCA
	RioBlanco	DB 007	TTTGA-				Y - C T A C T	CGCAGG	GGCCC	GTAT-NGC	CGGCR-		CCTCCA
	Rioblanco	DB 008	TTTGA-	K C I I G G	I G G G A G -		Y - C T A C T	CGCAGG	GGCCC	G A I - N G C	CGGCR-		CCTCCA
		DB 009		T W K Y U R	Y S R G A G -			CACIGI	GATGT	G - I C I G G I	CIIGGC	- I A C C I	CICIAG
	<b>B</b> 1	DB 010	TTTGA-	K C I I G G	I G G G A G -		Y - C T A C T	CGCAGG	GGCCC	GTAT-NGC	CGGCR-	CICICC	CCTCCA
)19	RILs	DB 011	TTTGA-				Y - C T A C T	CGCAGG		GTAT-NGC	CGGCR-		CCTCCA
		DB 012	TTTGA-	K C T T G G	T G G G A G -	TCTRATCC	Y - C T A C T	CGCAGG	GGCCC	GTAT-NGC	CGGCR-	CTCTCC	CCTCCA
		DB 013	TTTGA-	K C T T G G	TGGGAG-	TCTRATCC	Y - C T A C T	CGCAGG	GGCCC	GTAT-NGC	CGGCR-	CTCTCC	CCTCCA
)27	S Landraces	DB 014	WKYRR -	TWKYOR	YSRGAG-	TTCGGCCC	C - C T A C T	CACTGT	GATGT	G - T C T G G T	CTTGGC	- TACCT	CTCTAG
)21	5 Lanuraces	DB 015	WKYRR -	TWKYOR	YSRGAG-	TTCGGCCC	C - C T A C T	CACTGT	GATGT	G - T C T G G T	CTTGGC	- TACCT	CTCTAG
		DB 016	TTTGA-	KCTTGG	T G G G A G -	TCTRATCC		CGCAGG	GGCCC	GTAT-NGC	C G G C R -	CTCTCC	CCTCCA
		DB 017	WKYRR -	KMTYOR	YSRGAG-	TYYRRYCC		CRCWGK	GRYSY	GTWYOGGY	CKKSR0	OTMYCY	CYYYMR
	Bt0	DB 018	TTTGA-	KCTTGG	TGGGAG-	TCTRATCC	Y - C T A C T	CGCAGG	GGCCC	GTAT-NGC	CGGCR -	CTCTCC	CCTCCA
	Bto	DB 019	WKYRR-	TWKYOR	YSRGAG-	TTCGGCCC	C - C T A C T	SRYTKK	SAKGY	R - WYTSRT	MKKSGC	- TACCT	CTCTAG
		DB 020	WKYRRO	TWKYOR	Y S R R W K 0	KYCGRCYS	COYKMSW	ISRYTKK	SAKGY	R - WYTSRT	MKKSGC	- YAYYY	CYYYRR
		DB 021	TTTGGT	TTTTGA	CCAATTA	GCCGACTG	CTTGCGA	GGTTTG	CAGGC	A - A T T C A T	AGGCGC	- CATTC	CCTCGA
	Bt12	DB 022	TTTGRO	TYTTGR	Y S R R W K 0	KCYGAYYS	COYKMSW	ISGYWKG	SRSSC	ROATOCRY	MGGCR0	0 Y M T Y C	CCTCSA
		DB 023	TTTGRO	TYTTGR	YSRRWK0	KCYGAYYS	COYKMSW	ISGYWKG	SRSSC	ROATOCRY	MGGCG0	0 Y M T Y C	Y S T C S A
		DB 024	TTTGRO	TYTTGR	YSRRWK0	KCYGAYYS	COYKMSW	ISGYWKG	SRSSC	ROATOCRY	MGGCR0	OYMTYC	CCTCSA
	Bt13	DB 025	WKYRRO	TWKYOR	Y S R R W K 0	KYCGRCYS	COYKMSW	ISRYTKK	SAKGY	R - WYTSRT	MKKSGC	- Y A Y Y Y	CYYYRR
	DUIS	DB 026	TTTGGT	TTTGA	CCAATTA	GCCGACTG	CTTGCGA	GGTTTG	CAGGC	A - A T T C A T	AGGCGC	- CATTC	CCTCGA
		DB 027	WKYRRO	TWKYOR	Y S R R W K 0	KYCGRCYS	COYKMSW	ISRYTKK	SAKGY	R - W Y T S R T	MKKSGC	- Y A Y Y Y	CYYYRR
		DB 028	TTTGA-	TTGTGG	TGGGAG-	TTCGGCCC	C - C T A C T	CACTGT	GATGT	G - T C T G G T	CTTGGC	- TACCT	CTCTAG
		DB 029	TTTGGT	TTTGA	CCAATTA	GCCGACTG	CTTGCGA	GGTTTG	CAGGC	A - A T T C A T	AGGCGC	- CATTC	CCTCGA
		DB 030	TTTGGT	TTTGA	CCAATTA	GCCGACTG	CTTGCGA	GGTTTG	CAGGC	A - A T T C A T	AGGCGC	- CATTC	CCTCGA
		Undetermined	WKYRRO	K Y K Y O R	YSRRWKO	KYYRRYYS	YOYKMSW	ISRYWKK	SRSY	ROWYOCRY	MKKSR0	0 Y M Y Y Y	CYYYSR

Code	Genotype
DB-001	IDO444
DB-002	RioBlanco
DB-003 to -019	RILs
DB-020 to -027	S Landraces
DB-028	Bt0
DB-029	Bt12
DB-030	Bt13
DD-030	БПЭ

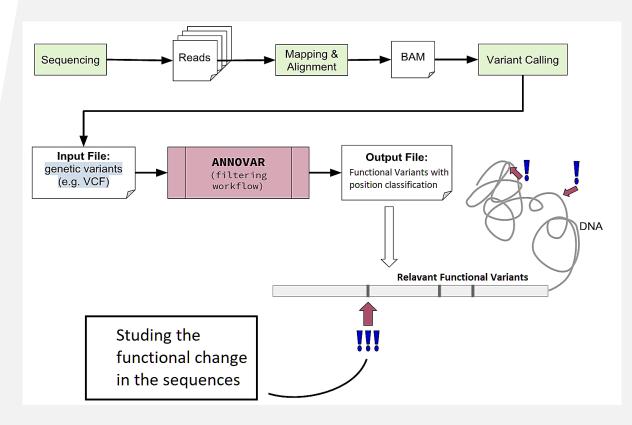


#### **FILTERED VARIANTS**



## IDENTIFYING FUNCTIONAL VARIANTS (ANNOVAR)

57





#### **POSITION AND FUNCTION-BASED CLASSIFICATION OF VARIANTS**

Position Type	No of variants
intergenic	3111
intronic	115
exonic	444
UTR3	170
downstream	232
upstream	185
UTR5	8
ncRNA_exonic	2
All	4267

Mutation Type	Exonic variants
frameshift Indels	14
stopgain	4
synonymous SNV	179
nonsynonymous SNV	235
nonframeshift Indels	12
All	444

# I

### **GENES WITH EXONIC VARIANTS**

- 171 annotated genes spanning 145 genetic loci found in 3-11 Mb region on 7DS
- Only 29 genes showed variation in the exonic regions in IDO444
- These genes showed wide variation in the proteins encoded
  - Receptor Kinases
  - F-box proteins
  - NB-LRRS
  - Rp1-like proteins
  - Cytochrome P450
  - Ring finger like proteins



### **GENES (29) WITH EXONIC VARIANTS**

Variant Type	Color Code				
stopgain	green				
	J				
synonymous SNV	red				
frameshift Indel	Blue				
Non Syn SNV	Black				

GeneID	Function	Variants	NonSyn Variants
TraesCS7D02G007400	NBS-LRR disease resistance protein-like	1	1
TraesCS7D02G008800	Acid invertase 1	1	1
TraesCS7D02G011600	Acetyltransferase component of pyruvate dehydrogenase complex	1	0
TraesCS7D02G011700	Protein arginine N-methyltransferase 7	1	0
TraesCS7D02G012300	Zn-dependent exopeptidases superfamily protein	5	4
TraesCS7D02G012400	receptor kinase 1	19	13
TraesCS7D02G012900	Rp1-like protein	9	5
TraesCS7D02G013200	receptor kinase 1	39	17
TraesCS7D02G013300	Rp1-like protein	48	31
TraesCS7D02G013400	disease resistance family protein / LRR family protein	47	28
TraesCS7D02G017600	NBS-LRR-like resistance protein	25	11
TraesCS7D02G017700	NBS-LRR-like resistance protein	46	32
TraesCS7D02G017900	NBS-LRR disease resistance protein-like protein	32	16
TraesCS7D02G019200	Glycosyltransferase	2	0
TraesCS7D02G020300	F-box protein	18	9
TraesCS7D02G020400	Benzyl alcohol O-benzoyltransferase	3	1
TraesCS7D02G021100	BURP domain protein RD22	16	11
TraesCS7D02G021700	F-box protein	6	3
TraesCS7D02G022000	receptor-like protein kinase 1	17	6
TraesCS7D02G022100	Disease resistance protein (NBS-LRR class) family	27	16
TraesCS7D02G022200	Protein kinase family protein	6	5
TraesCS7D02G022300	Disease resistance protein (NBS-LRR class) family	4	1
TraesCS7D02G022500	Protein FRA10AC1	1	1
TraesCS7D02G022600	RING finger protein	1	1
TraesCS7D02G022900	Invertase inhibitor	2	1
TraesCS7D02G023000	RING finger protein	9	4
TraesCS7D02G023100	RING finger family protein	2	1
TraesCS7D02G023300	Cytochrome P450	21	8
TraesCS7D02G023400	F-box protein	35	26

#### **POTENTIAL TARGETS IN NINE GENES**

ID	IDO444	RioBlanco	Pos	Mutation	Gene	Function	Mutation NT	Mutation AA	Mutant domains	GeneStart	GeneStop	Size(kb)
S7D_4268568	AG	А	exon3	frameshift Indel	TraesCS7D02G008800	Acid invertase 1	c.1369delC	p.L457fs	C-terminal Transmembrane?	4267640	4270860	3.22
\$7D_5331386	G	GTT	exon3	frameshift Indel	TraesCS7D02G012400	receptor kinase 1	c.211_212insTT	p.G71fs	Before binding site of protein kinase	5330809	5337313	6.504
S7D_5331411	ACG	А	exon3	frameshift Indel	TraesCS7D02G012400		c.237_238del	p.D79fs	Before binding site of protein kinase	5330809	5337313	6.504
\$7D_5512644	AAATAG	А	exon8	frameshift Indel	TraesCS7D02G013200	receptor kinase 1	c.3939_3943del	p.F1313fs	C-terminal non domain region	5510346	5518574	8.228
\$7D_5512713	AT	А	exon8	frameshift Indel	TraesCS7D02G013200		c.3874delA	p.I1292fs	C-terminal non domain region	5510346	5518574	8.228
\$7D_5634198	G	А	exon1	stopgain	TraesCS7D02G013300	Rp1-like protein	c.C2602T	p.Q868X	Terminal end (C) of protein	5634184	5636799	2.615
\$7D_5648558	тс	Т	exon1	frameshift Indel	TraesCS7D02G013400	disease resistance family protein	c.49delG	p.E17fs	N-terminal before domain	5645812	5648607	2.795
S7D_7750174	т	С	exon2	stopgain	TraesCS7D02G017700	NBS-LRR-like resistance protein	c.C106T	p.R36X	in Rx N-terminal domain	7749593	7754061	4.468
S7D_7750708	С	т	exon2	stopgain	TraesCS7D02G017700		c.C640T	p.Q214X	in NB-ARC domain	7749593	7754061	4.468
S7D_7752652	A	AGCTTGGT AACGTTCC CGAT	exon2	frameshift Indel	TraesCS7D02G017700		c.2585_2603del	p.S862fs	after NB-ARC domain	7749593	7754061	4.468
S7D_7752984	CACAA	С	exon2	frameshift Indel	TraesCS7D02G017700		c.2916_2917insACAA	p.V972fs	after NB-ARC domain	7749593	7754061	4.468
S7D_10478651	тс	т	exon2	frameshift Indel	TraesCS7D02G022100	Disease resistance protein (NBS-LRR class) family	c.919delG	p.E307fs	after NB-ARC domain	10477394	10480078	2.684
S7D_10478655	т	TGGTA	exon2	frameshift Indel	TraesCS7D02G022100		c.915_916insTACC	p.1306fs	after NB-ARC domain	10477394	10480078	2.684
S7D_10478659	CG	С	exon2	frameshift Indel	TraesCS7D02G022100		c.911delC	p.S304fs	after NB-ARC domain	10477394	10480078	2.684
S7D_10478787	т	А	exon2	stopgain	TraesCS7D02G022100		c.A784T	p.R262X	after NB-ARC domain	10477394	10480078	2.684
S7D_10869822	TA	т	exon1	frameshift Indel	TraesCS7D02G023300	Cytochrome P450	c.512_513insT	p.A171fs	after cytochrome domain transmembrane?	10868164	10870414	2.25
S7D_10963456	т	TGC	exon1	frameshift Indel	TraesCS7D02G023400	F-box protein	c.203_204del	p.R68fs	in F box domain	10962062	10963681	1.619



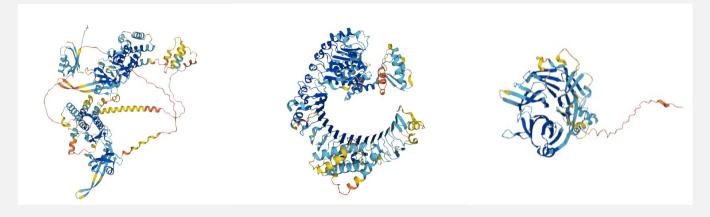
# ONGOING PROJECTS – USDA-NIFA AFRI Validation of the candidate genes

- Three genes (primary targets) being cloned into vectors
- Transformation of RioBlanco with cloned genes
- Studying gain of function in transformed RioBlanco
- Studying loss of function using Crisper/Cas9 mediated transformation
- > Studying loss of function using EMS mutation lines



#### ONGOING PROJECTS – USDA-NIFA AFRI Validation of the candidate genes

- Sequencing the three candidate genes
- Developing in-situ gene models to compare the structural variations in the protein folding of the candidate genes which, in turn, would reveal the mode of action of the genes in imparting resistance to the bunt diseases.





#### **ONGOING PROJECTS – USDA-NIFA AFRI**

> Fine mapping of the two major QTL, *QDB.ui-6DL* 

and QDB.ui-7D

- Developing genomic selection model to select major and minor genes
- Implementing marker assisted selection to pyramid multiple resistance genes
- Releasing bunt resistance cultivars



## **COLLABORATIVE TEAM**





















#### ACKNOWLEDGEMENT





University of Idaho

Idaho Agricultural Experiment Station