

Wheat ~~Exome~~ Promotome Capture

A sub-contractor project of WGIN 3

Michael Hammond-Kosack (WGIN PA)



MYcroarray

Ann Arbor, Michigan, USA

MYbaits[®]

Target Enrichment Kits for NGS

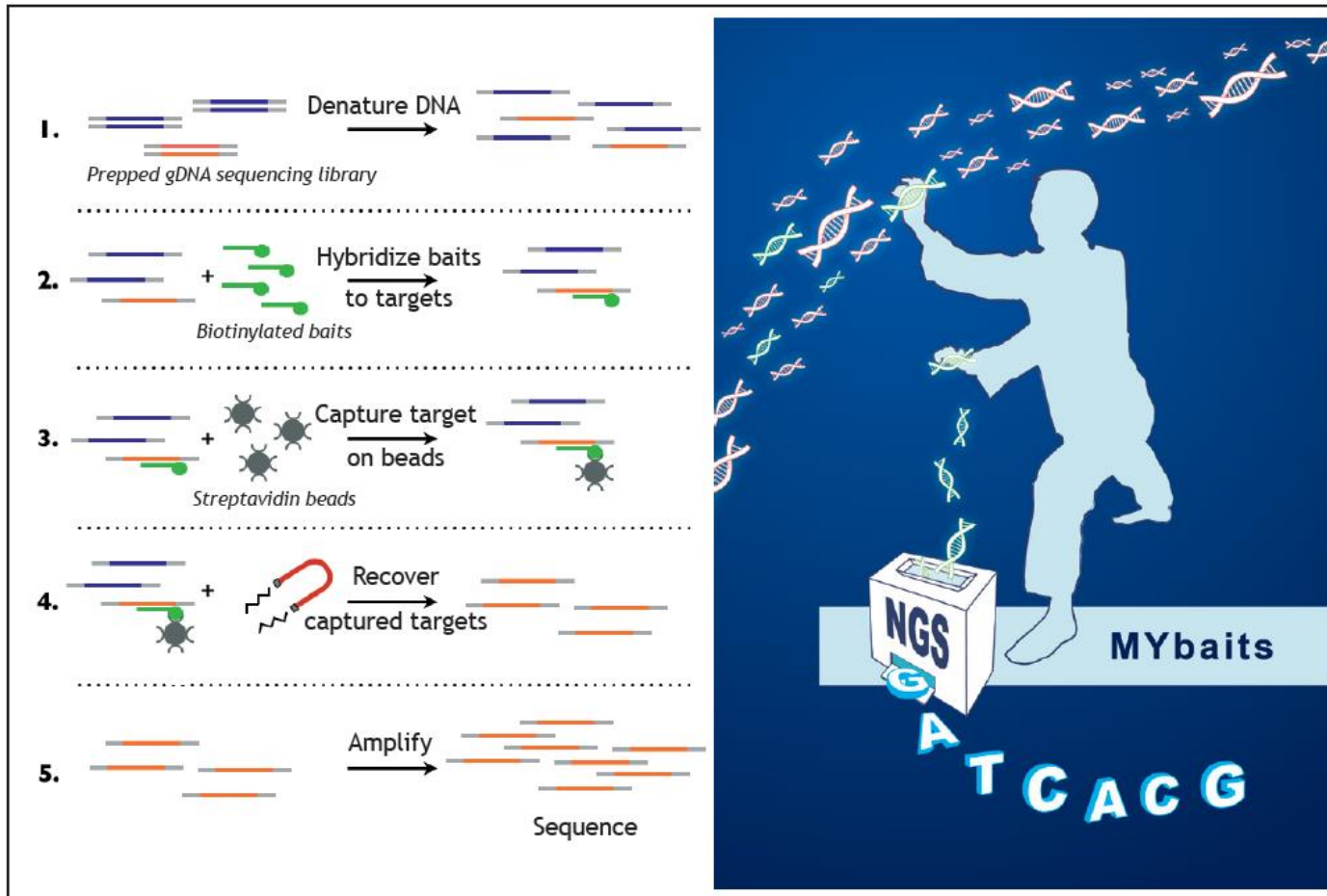
Any locus. **Any** organism.

*Simultaneously enrich thousands of
custom loci for ultra-efficient NGS*



Custom bait libraries for targeted sequencing

Mybaits is a fully customisable liquid-phase DNA capture system for targeted sequencing



High percentage of reads on target.

Complete workflow:



Summary of exome capture discussions @ WGIN Stakeholder meeting 16th April 2015

- Focus : promoter sequences (~1kb) – **NOVELTY**

“Exome Capture”  “Promotome Capture”

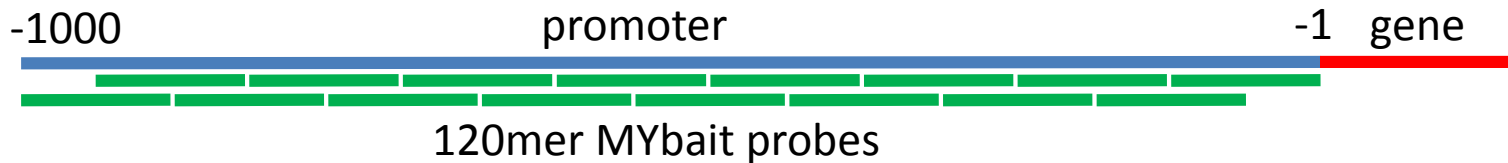
- A, B and D genome sequences to be captured individually.

nb: The Bioinformatics will be quite challenging to ID the 3 homoeologous promoters
[thx to Rob and Keywan in advance]

Funding & Design

- Defra's WGIN 3 funding for this sub-contractor project allowed for the use of **20,000 MYbaits probes** and **96 "Reactions"** (which equates to **96 Wheat Cultivars genomic DNA**)

- Every **1000bp of promoter** sequence/gene requires **16 overlapping probes** for twofold coverage:



- 3 wheat homoeologues (A, B, D) ➡ **48 Mybait probes / gene.**

- Thus Defra's funding will allow for the capture of **416 promoter sequences.**

- The WGIN Management Team decided on **10 Trait Categories**, and thus **40 promoter sequences per trait**

Trait Categories & Trait Coordinators

1. **Yield Resilience** –
Cristobal Uauy (JIC), Elizabete Carmo-Silva (Lancs U)
2. **Grain Composition** –
Peter Shewry/Rowan Mitchell (RRes), Kay Trafford (NIAB)
3. **Grain Development** – Cristobal Uauy (JIC)
4. **Biotic Stress (fungi and insects)** –
Matthew Moscou (SL) & Kim Hammond-Kosack (RRes)
5. **Abiotic Stress (drought, high temp)** –
John Foulkes (UoN), Eric Ober (NIAB)
6. **Nutrient Use Efficiency** –
Malcolm Hawkesford (RRes), Martin Broadley (UoN)
7. **Canopy Development / Whole Plant Architecture** –
Alison Bentley (NIAB), Simon Griffiths (JIC)
8. **Flower biology** – Zoe Wilson (UoN), Scott Bowden (JIC)
9. **Root architecture** – Malcolm Bennett (UoN), Peter Buchner (RRes)
10. **Recombination** – James Higgins (Leicester U)

The 96 Wheat Cultivars

1. All Wheat Management Team members and ALL Trait coordinators were asked for their preferred cultivars



2. Eventually 143 Cultivars were put forward.



3. Reduced to 96 cultivars after considering

- use in current and previous WGIN experiments
- Number of nominees
- Trait coverage
- Individual discussions with nominees (but no bribes)

	Cultivar	nominator(s)	1. Yield resilience
1	A. speltoides 2140022	LS/GA	
2	Abbot	SB	
3	Aegilops peregrina (variabil	KT	
4	Alcedo	RAGT	
5	Ambrosia	SB	
6	Avalon	SG/MH/VM/SB/CL/EO+JF	
7	Badger	CL	
8	Bobwhite	KK	
9	Brompton	LG	
10	Buster	SB	
11	Cadenza	SG/MH/VM/SB/CL	
12	Cellule	RAGT	
13	Charger	SB	
14	Chinese Spring	SG	
15	Claire	RAGT/SG/PS/MH/CL	
16	Coppadra	KK	
17	Cordiale	PS/MH	
18	Cougar	RAGT	
19	Crusoe	RAGT/MH	RL yield control, introgressio
20	Dickens	LG	
21	Einstein	SB	
22	ENT-228	LS/GA	
23	Fielder	KK	
24	Flanders	RAGT	

25	Gallant	MH	
26	Garcia	SG/CL	
27	Gatsby	ECS	low Rubisco
28	Gladiator	SB	
29	Graham	LG/syn	High yield
30	Hereford	KHK/VM	Highest yielder in RL trials, 2
31	Hereward	RAGT/PS/MH/SB/CL	
32	Hobbit	SB	
33	Hustler	SB	
34	Isengrain	PS	
35	Istabraq	PS/MH	
36	JB Diego	RAGT	Proven yield resilience over
37	Kronos	WGIN MM team	
38	KWS Santiago	RAGT	Proven yield resilience over
39	KWS Silverstone	LG	High yield
40	KWS Siskin	RAGT	High yield
41	KWS Trinity	LG	
42	Malacca	PS/MH/CL	
43	Maris Huntsman	SB	
44	Maris Widgeon	MH	
45	Marksman	PS	
46	Mercia	MH	
47	Napier	CL	
48	Oakley	RAGT	

49	Paragon	SG/MH/AR/VM	
50	Piko	RAGT	
51	Reflection	LG	High yield
52	Relay	RAGT	
53	Revelation	LG	
54	Rialto	RAGT/SB/PS	
55	Riband	MH	
56	Riband	KK	
57	Robigus	RAGT/SG/MH/AR	
58	Savannah	CL	
59	Scout	RAGT	
60	Sear Synthetic	SG	
61	Skyfall	RAGT	RL yield control
62	Soisson	RAGT/PS/MH	
63	Solstice	RAGT/GA_LS/MH	RL yield control
64	Spark	PS/CL	
65	Stigg	MH	
66	Sumai 3	RAGT	
67	<i>T. monoccocum</i> MDR031	KHK/VM	
68	<i>T. monoccocum</i> MDR037	LS/GA/VM	
69	<i>T. monoccocum</i> MDR043	KHK/VM	
70	<i>T. monoccocum</i> MDR045	LS/GA	
71	<i>T. monoccocum</i> MDR046	RRES/VM	
72	<i>T. monoccocum</i> MDR049	LS/GA	

73	<i>T. monoccocum</i> MDR308	KHK	
74	<i>T. monoccocum</i> MDR657	LS/GA	
75	Taichung 29	KK/JRudd	
76	Ukrainka	PS	
77	USU-Apogee	KK	
78	Valoris	PS	
79	Veranopolis	KK	
80	Watkins 115	LS/GA	
81	Watkins 141	SG	
82	Watkins 160	SG	
83	Watkins 199	LS/GA/SG	
84	Watkins 203	VM	
85	Watkins 209	SG	
86	Watkins 246	SG	
87	Watkins 292	SG	
88	Watkins 387	SG	
89	Watkins 579	LS/GA	
90	Watkins 624	LS/GA	
91	Watkins 733	VM	
92	Watkins 777	VM/SG	
93	Watkins 786	VM	
94	Xi19	RAGT/PS/MH	
95	Yumai 34	PS	
96	Zebedee	EO+JF	

The Gene Lists

Every gene list spreadsheet contained the following columns:

- A) gene list
- B) Gene designation from literature
- C) Gene function or product
- D) AA genome (add ENSEMBL no.)
- E) BB genome (add ENSEMBL no.)
- F) DD genome (add ENSEMBL no.)
- G) unknown genome location (add ENSEMBL no.)
- H) Key cultivars/lines (maximum of 2)
- I) Doi for key publication.
- J) comments
- K) nominator
- L) email address

An example – Root Architecture (Trait 9)

A	B	C	D	E
gene list	Gene designation from lit	Gene function or product	AA genome (add ENSEMBL no.)	BB genome (add ENSEMBL no.)
1	none	GATB-Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase	TRIAE_CS42_2AL_TGACv1_097369_AA0323960	TRIAE_CS42_2BL_TGACv1_129918_AA0399390
2	none	ARF7, encodes auxin-regulated transcriptional activator	TRIAE_CS42_7AS_TGACv1_570721_AA1839790	none
3	none	Galactoside 2-alpha-L-fucosyltransferase [Source: UniProtKB/Swiss-Prot]	TRIAE_CS42_6AL_TGACv1_472947_AA1527380	TRIAE_CS42_6BL_TGACv1_501779_AA1620700
4	none	TAA1, TAA1 is involved in the shade-induced production of auxin	TRIAE_CS42_1AS_TGACv1_019443_AA0066620	
5	none	PREDICTED: auxin-responsive protein SAUR36-like	TRIAE_CS42_2AL_TGACv1_096574_AA0319940	TRIAE_CS42_2BL_TGACv1_133232_AA0441960
6	none	DAO1, IAA oxidases expressed in many different plant species	TRIAE_CS42_2AL_TGACv1_095532_AA0311890	TRIAE_CS42_2BL_TGACv1_132715_AA0439330
7	none	Pectin lyase-like superfamily protein [Source: UniProtKB/Swiss-Prot]	TRIAE_CS42_2AL_TGACv1_095550_AA0312050	TRIAE_CS42_2BL_TGACv1_130085_AA0403410
8	none	IAA14, is a member of the Aux/IAA protein family. Involved in auxin signaling	TRIAE_CS42_5AL_TGACv1_374359_AA1197870	TRIAE_CS42_5BL_TGACv1_405745_AA1334340
9	none	Domain of unknown function DUF547 (IPR006869), conserved domain	TRIAE_CS42_7AS_TGACv1_571319_AA1846790	TRIAE_CS42_7BS_TGACv1_593096_AA1948130
10	none	PIN2, Encodes an auxin efflux carrier that is similar to PIN1	TRIAE_CS42_7AL_TGACv1_557950_AA1788290	TRIAE_CS42_7BL_TGACv1_578580_AA1897450
11	none	Protein BONZAI 3 [Source: UniProtKB/TrEMBL; accession: A0A017770.1]	TRIAE_CS42_1AL_TGACv1_000719_AA0017770	TRIAE_CS42_1BL_TGACv1_031297_AA0111050
12	none	AUX1, Encodes an auxin influx transporter. AUX1 responsible for auxin uptake	TRIAE_CS42_1AL_TGACv1_001450_AA0030760	TRIAE_CS42_1BL_TGACv1_031778_AA0120460
13	none	QWRF family (IPR007573), The plant auxin receptor	TRIAE_CS42_5AL_TGACv1_375923_AA1228910	TRIAE_CS42_5BL_TGACv1_406012_AA1339030
14	none	LAX3, Encodes an auxin influx carrier LAX3 (Like AUX1)	TRIAE_CS42_4AL_TGACv1_291302_AA0993610	TRIAE_CS42_4BS_TGACv1_329047_AA1097390

F	G	H	I
DD genome (add ENSEMBL no.)	unknown genome location (add ENSEMBL no.)	Key cultivars/lines (maximum of 2)	Doi for key publication.
TRIAE_CS42_2DL_TGACv1_160509_AA0551270		no relevant cultivars/lines	Qin et al., Front Plant Sci.
TRIAE_CS42_7DS_TGACv1_622375_AA2038600		no relevant cultivars/lines	Wilmoth et al., 2005, Plant
TRIAE_CS42_6DL_TGACv1_527012_AA1697180		no relevant cultivars/lines	none
TRIAE_CS42_1DS_TGACv1_081721_AA0262410		no relevant cultivars/lines	Stepanova et al., 2008, C
TRIAE_CS42_5DS_TGACv1_457137_AA1482860		no relevant cultivars/lines	none
TRIAE_CS42_2DL_TGACv1_158281_AA0514640		no relevant cultivars/lines	Porco et al., 2016, PNAS,
TRIAE_CS42_2DL_TGACv1_159273_AA0535750		no relevant cultivars/lines	none
TRIAE_CS42_5DL_TGACv1_435293_AA1448500		no relevant cultivars/lines	Fukaki et al., 2002, Plant
TRIAE_CS42_7DS_TGACv1_621528_AA2018470		no relevant cultivars/lines	none
TRIAE_CS42_7DL_TGACv1_602827_AA1969450		no relevant cultivars/lines	DOI: 10.1007/978-1-4939
TRIAE_CS42_1DL_TGACv1_061939_AA0205980		no relevant cultivars/lines	none
TRIAE_CS42_1DL_TGACv1_061187_AA0188170		no relevant cultivars/lines	DOI: 10.1104/pp.15.0090
TRIAE_CS42_5DL_TGACv1_433963_AA1426330		no relevant cultivars/lines	none
none		no relevant cultivars/lines	Swarup et al., 2008, Natu

J

K

L

comments	nominator	email address
Orthologous to LOC_Os11g34210-OsGatB, the Subunit of tRNA-Dependent Amidotransferase, Is Required for Primary Root Development in Rice	Peter Buchner	peter.buchner@rothamsted.ac.uk
ARF7 and ARF19 required for auxin response to develop lateral roots and root hairs (root branching)	Malcolm Bennett	malcolm.bennett@nottingham.ac.uk
	Peter Buchner	peter.buchner@rothamsted.ac.uk
Orthologous to Arabidopsis TAR2, a protein with similarity to the TAA1 tryptophan aminotransferase involved in IAA biosynthesis. TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant root development.	Malcolm Bennett	malcolm.bennett@nottingham.ac.uk
It has been shown that auxin-responsive proteins are involved in root morphology	Peter Buchner	peter.buchner@rothamsted.ac.uk
DAO1 specifically expressed in trichoblast cell files and controls root hair elongation	Malcolm Bennett	malcolm.bennett@nottingham.ac.uk
	Peter Buchner	peter.buchner@rothamsted.ac.uk
SLR/IAA14 is a key regulator in auxin-regulated growth and development, particularly in lateral root formation.	Malcolm Bennett	malcolm.bennett@nottingham.ac.uk
	Peter Buchner	peter.buchner@rothamsted.ac.uk
Required for lateral root and root hair development. Also involved in root gravitropic response.	Malcolm Bennett	malcolm.bennett@nottingham.ac.uk
In Arabidopsis the copine family consist of BON1 (BONZAI 1), BON2 (BONZAI 2) and BON3 (BONZAI 3). The copine family in plants may have effects in promoting growth and development in addition to repressing cell death [PMID: 11544183, PMID: 16367962].	Peter Buchner	peter.buchner@rothamsted.ac.uk
Plays a role in lateral root initiation and positioning	Malcolm Bennett	malcolm.bennett@nottingham.ac.uk
The plant augmin complex is involved in assembly of microtubules (MT) arrays during mitosis and contains eight subunits (AUG1 -AUG8). AUG8 belongs to the plant QWRF motif-containing protein family, which also includes microtubule-associated protein ENDOSPERM DEFECTIVE 1 [PMID: 19151224, PMID: 22535409] and SNOWY COTYLEDON 3 [PMID: 20978221]. AUG8 binds the microtubule plus-end and participates in the reorientation of microtubules in hypocotyls (the stem of a germinating seedling) [PMID: 23735294, PMID: 23749846].	Peter Buchner	peter.buchner@rothamsted.ac.uk
Promotes lateral root emergence	Malcolm Bennett	malcolm.bennett@nottingham.ac.uk

Dataset
Triticum aestivum genes (TGACv1)

Filters
Gene stable ID(s): [ID-list specified]

Attributes
Gene stable ID
Transcript stable ID

Dataset
[None Selected]

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

REGION:

GENE:

Limit to genes ... with ENA/GenBank ID(s) Only Excluded

ID list limit

Gene stable ID(s)

- TRIAE_CS42_2AL_TGACv1_097369_AA0323960
- TRIAE_CS42_2BL_TGACv1_129918_AA0399390
- TRIAE_CS42_2DL_TGACv1_160509_AA0551270
- TRIAE_CS42_7AS_TGACv1_570721_AA1839790
- none
- TRIAE_CS42_7DS_TGACv1_622375_AA2038600
- TRIAE_CS42_6AL_TGACv1_472947_AA1527380
- TRIAE_CS42_6BL_TGACv1_501779_AA1620700
- TRIAE_CS42_6DL_TGACv1_527012_AA1697180
- TRIAE_CS42_1AS_TGACv1_019443_AA0066620
- TRIAE_CS42_1DS_TGACv1_081721_AA0262410
- TRIAE_CS42_2AL_TGACv1_096574_AA0319940
- TRIAE_CS42_2BL_TGACv1_133232_AA0441960
- TRIAE_CS42_5DS_TGACv1_457137_AA1482860
- TRIAE_CS42_2AL_TGACv1_095532_AA0311890
- TRIAE_CS42_2BL_TGACv1_132715_AA0439330
- TRIAE_CS42_2DL_TGACv1_158281_AA0514640

Dataset
Triticum aestivum genes (TGACv1)

Filters
Gene stable ID(s): [ID-list specified]

Attributes
Gene stable ID
Transcript stable ID
Flank (Gene)
Upstream flank [1000]

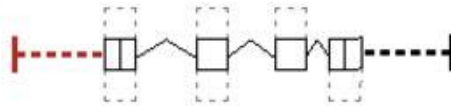
Dataset
[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

- Features
- Homologs
- Structures
- Sequences
- Variation

☐ SEQUENCES:

Sequences (max 1)



- | | |
|--|---------------------------------------|
| <input type="radio"/> Unspliced (Transcript) | <input type="radio"/> 5' UTR |
| <input type="radio"/> Unspliced (Gene) | <input type="radio"/> 3' UTR |
| <input type="radio"/> Flank (Transcript) | <input type="radio"/> Exon sequences |
| <input checked="" type="radio"/> Flank (Gene) | <input type="radio"/> cDNA sequences |
| <input type="radio"/> Flank-coding region (Transcript) | <input type="radio"/> Coding sequence |
| <input type="radio"/> Flank-coding region (Gene) | <input type="radio"/> Peptide |

Upstream flank

Upstream flank

Downstream flank

Downstream flank



Dataset

Triticum aestivum genes (TGACv1)

Filters

Gene stable ID(s): [ID-list specified]

Attributes

Gene stable ID
Transcript stable ID
Flank (Gene)
Upstream flank [1000]

Dataset

[None Selected]

Export all results to Unique results only

Email notification to

View rows as Unique results only

```
>TRIAE_CS42_1AL_TGACv1_001450_AA0030760|TRIAE_CS42_1AL_TGACv1_001450_AA0030760.2;T
AGAACAAGTCCAATCATACTTAACCAGCAATGGGAAGGACCATTGGAGCCCAACCTCTGT
GCTTGTGGAGATCTCAGCTGCCACTTGCAGCAGAAGCCTTGCCCCTCTGACTTGCACCC
CTCATTATTTTCGTTACTATAATAATTTTCAGAATTCTAACCAATAGTCAGATTCCGTAA
TTGCTTGAAGGATCATAGGGATAAATATTCTCGAAACATGGTTTATTTCTAGATTTCCAC
AAAGCTCACAACATCGCAACAACCTCTACTGCTATATTCCCCACTAGCTTTTCTAGGAAC
ACCAAAAATCCAACCACAAAAGTCATGTACATATCCTACATTCATATCTATCCCGAGAGC
ACACTTGACCACATACCTCAAGCATAACTAGGCCAAAGAGCAATGGAAAAATACATGATCT
AATGTTTCTTTCTCATTACAAAATCTGCACTTATCACTACCAAACCATCCTCTAGCAAAC
AAGCTATCTTTTGTGAGTATACTATGCCCAAACATAATCCAAAGAAAGACCTTAAATGTC
AACGGAACTTTAAATAAACCAAAGCTTATTATAAGGACAATCTACCCGTATAACCTCAAC
ATATTATATAAAGATTTTACCAAGAAGTTACCAATGTATAAAAATGGCAGAGATATAGATG
GATATAAGCGTATGAAAGATAGGAATTGACACACCATAGCAGAATAATAAAAATGTGCCAA
CATATAAAAAACATGAATCAAGGCCAAAACCATTAATTCGGTTCATAATATGGTGTATAT
CCTATTTATTGTATTTATTTACTATTACCTATTGATTCTCTACAATTTTAGTTTTCCAC
CGTGTGATCGTTGTTGTTTATATTTGTAGATTTCTTTTAGTACCACGCTAATCGAATAA
ATCGTCCAAGATGGAATACATTTGATGCGAAAGCCTGCCAAAAGCATTATTACCATAGTC
AAGTTCTATAAAAATCCACACATTACCTGTACCCTATTTCG
>TRIAE_CS42_1AS_TGACv1_019443_AA0066620|TRIAE_CS42_1AS_TGACv1_019443_AA0066620.1
CTAATAGGTTGGGGATACCACTGTCTACCTAACCACTCAACCATATATTTAGGTAAAAT
TGTCATCCTTACCGGCCACGCCGCCACGCATCGCACGAACTAACCCTAACCCTAAT
TCCATCCACCTATCCCGCCCAACCGCCCACTAAGCTCAGCCCGCCCTACATACAAACCA
```


Timeline

- 1) **December 2016:** obtain remaining gene-lists
- 2) **December 2016:** obtain seed for all 96 cultivars
- 3) **January 2017:** make chromosomal DNAs
- 4) **January 2017:** send FASTA file to Mycroarray for them to design Mybaits
- 5) **February 2017:** send chromosomal DNAs to Mycroarray to conduct the Promotome Capture Experiment.
- 6) **March 2017:** Next Generation Sequencing
- 7) **April 2017:** Sequences returned and analysed at RRes
- 8) **Summer 2017:** Data & Analysis released to Wheat Community

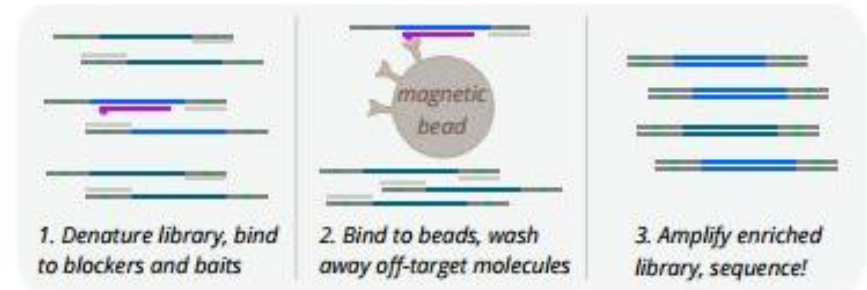
Rapid and easy-to-use

- Fully custom target enrichment for NGS
- Reliable, flexible, and affordable
- RNA baits and hybridization reagents in a convenient kit
- Couple with your preferred library preparation
- Target DNA sequences from animals, plants, & microbes
- Any sample source - even archival or ancient DNA
- Capture multiple taxa with the same kit
- Free bait design assistance



In-solution hybridization capture

- 1) Biotinylated RNA baits bind to targets
Library renaturation prevented with blockers
- 2) Hybrids isolated with magnetic beads
Non-target molecules washed away
- 3) Amplify enriched library with universal primers
= *Highly efficient targeted sequencing!*



Complete workflow:

