Agriculture and
Agri-Food Canada

Agriculture et
Agroalimentaire Canada

Canada

Haplotag:

your passport to anywhere in the oat world

Nick Tinker

Ottawa Research and Development Centre

Goal: EVERYONE understands……

1. What is Genotyping-by-Sequencing (GBS)
2. Briefly, how we use it in cultivated oat
3. **What we JUST DISCOVERED by applying GBS in wild oat species**

- **Everyone gets a GBS passport!**

Produced by new program "Haplotag"

Tinker et al. (2016)

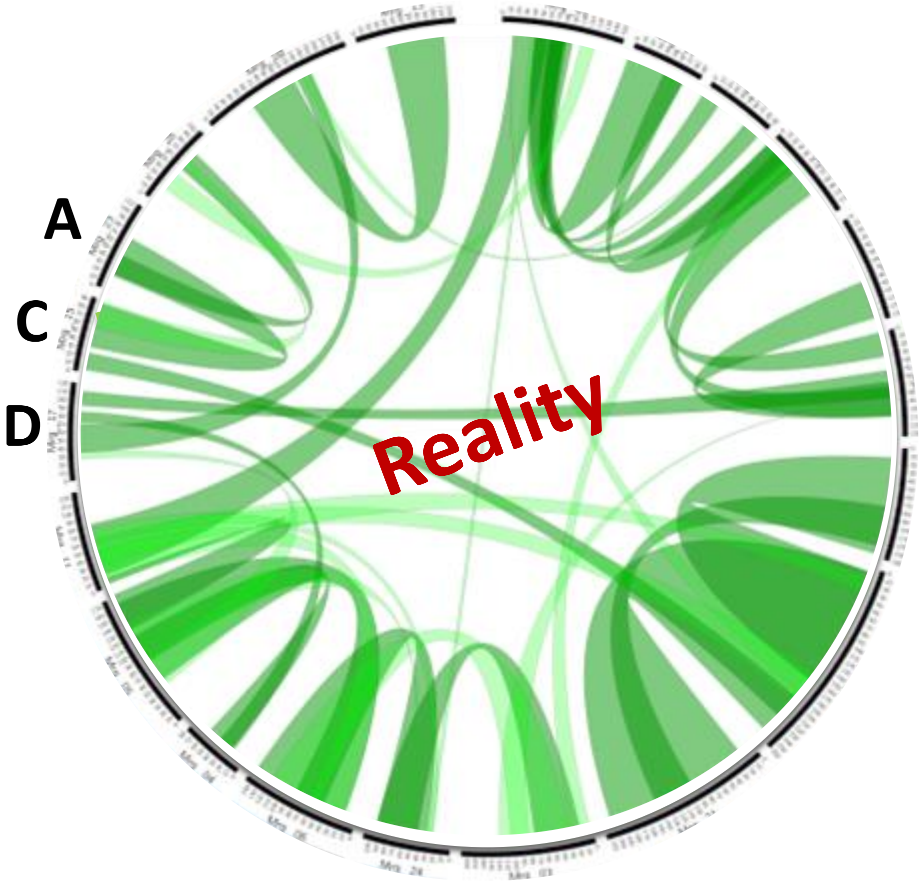
Genes, Genomes and Genetics

<u>Cultivated</u> Oat Genome – Why is it a challenge ?

- Large repetitive genome (13 billion bases)
- Not yet sequenced*
- Hexaploid (AA+CC+DD)
- Chromosome rearrangements

We have a lot to learn more about chromosome structure and origin.

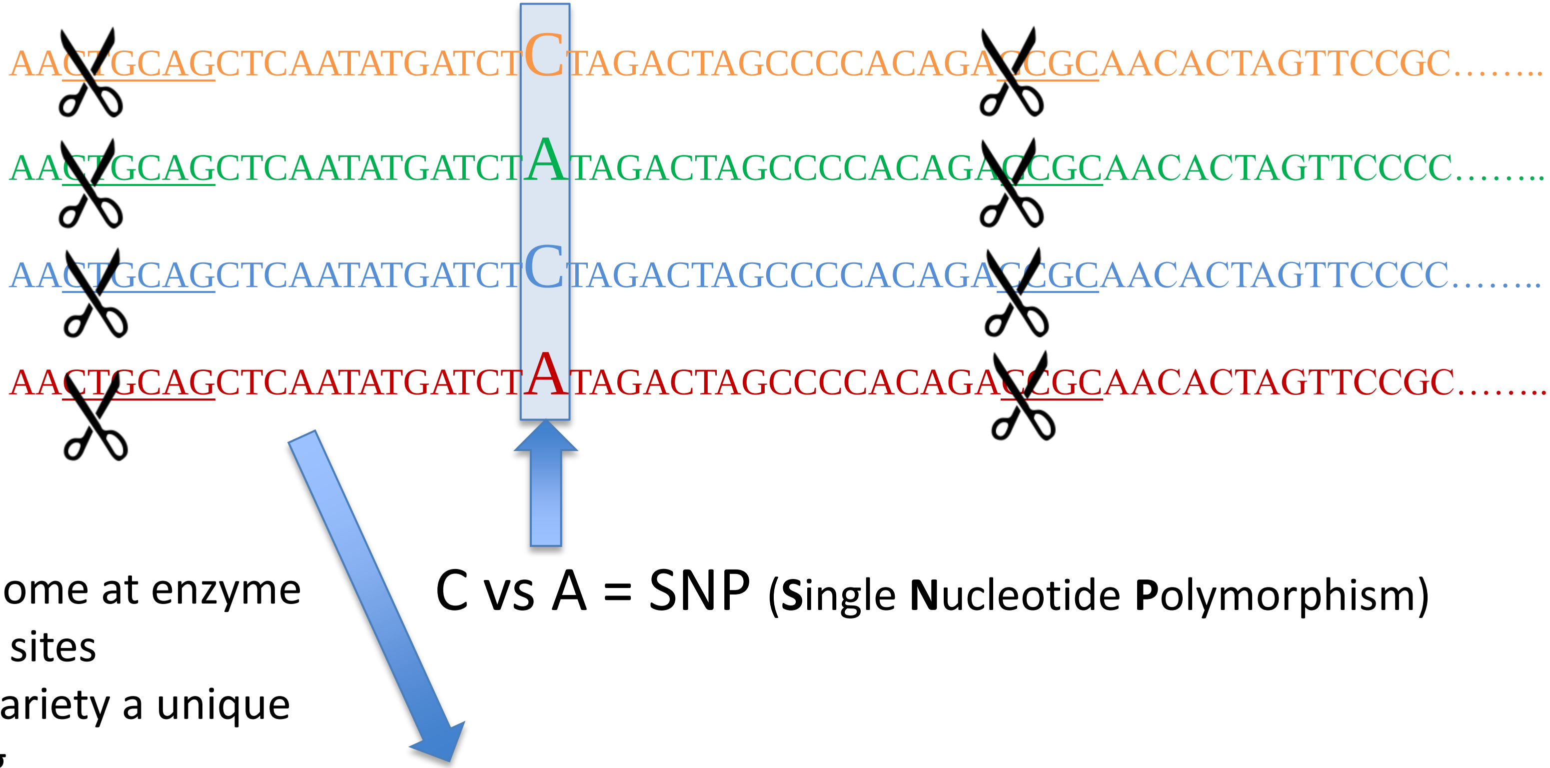
Future of oat could depend on finding and integrating alleles from A, C, D genome ancestors (?)



What is GBS, what does Haplotag do?

Imagine:

The complete 13-billion-base genomes of each variety……….



Variety-1: AACTGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCGCAACACTAGTTCCGC……..

Variety-2: AACTGCAGCTCAATATGATCTATAGACTAGCCCCACAGACCGCAACACTAGTTCCCC……..

Variety-3: AACTGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCGCAACACTAGTTCCCC……..

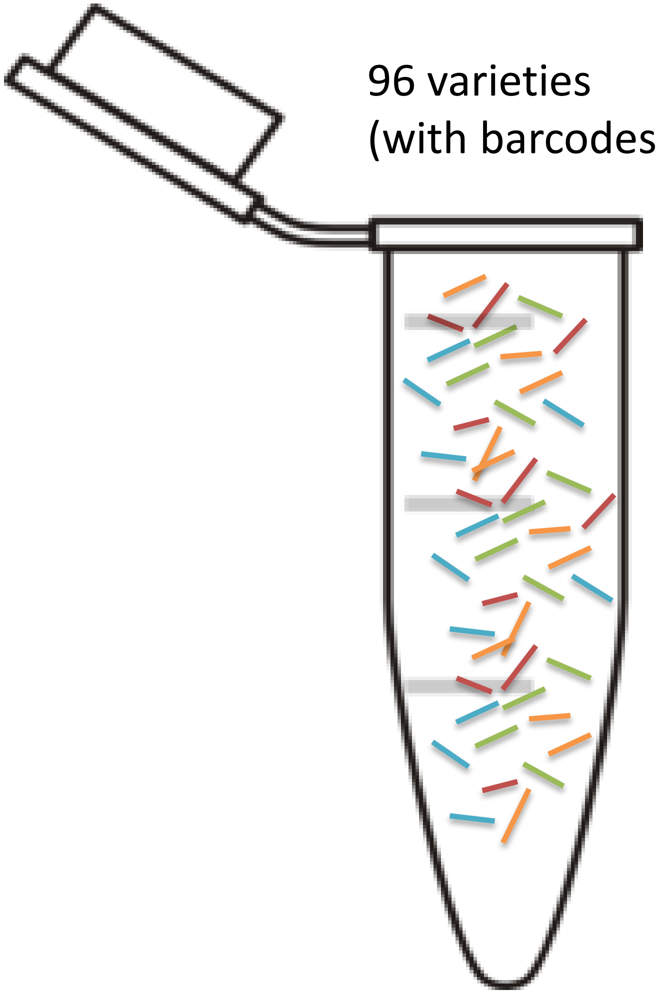
Variety-4: AACTGCAGCTCAATATGATCTATAGACTAGCCCCACAGACCGCAACACTAGTTCCGC……..

C vs A = SNP (**S**ingle **N**ucleotide **P**olymorphism)

- Cut the genome at enzyme recognition sites
- Give each variety a unique barcode tag

GGTACCT + TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGAC

Now we mix all the fragments to sequence them all at once:



96 varieties
(with barcodes)

400 Million reads
= "sequence tags"

```
TGCAGTGCGTGTGCCGTGGAAATTTCCGAGCCGACTGCTCGCCTGCTCCACTCCAGCAGGGTTA
TGCAGTGCGTGTGCCGTGGAAATTTCTGAGCCGACTGCTCGCCTGCTTCACACCGAGATCGGAA
TGCAGTGCGTGTGCGCAGCACAAAATCCTTTGTGTCTCTCGAGCTCTAGTGTTCATCCACCTAC
TGCAGTGCGTGTGGTGAGTCCCCGCCGTTCGTGGTCGCCCATGGAGGCTGGGCCTGATCTGGGC
TGCAGTGCGTGTGTCGCGGAAATTTCCGAGCCGACTGCTCGCCTGCTCCACGCAGAGATCGGAA
TGCAGTGCGTGTGTCGCGGAAATTTCCGAGCCGACTGCTCGCCTGCTCCACGCCGAGATCGGAA
TGCAGTGCGTGTGTCGCGGAAATTTCCGAGCCGACTGCTCGCCTGCTCCACGCTGAGATCGGAA
TGCAGTGCGTGTGTCGCGGAACTTTCCGAGCCGACTGCTCGCCTGCTCCACGCCGAGATCGGAA
TGCAGTGCGTGTTCAAGAACCTGGACGACGGCACGGTGTTCGTGGTCGACGAGGTGGGCGCGGA
TGCAGTGCGTGTTCAAGAACCTGGACGACGGCACGGTGTTCGTTGTCGACGAGGTGGGCGCGGA
TGCAGTGCGTGTTCCCCCGTCACGTTGAGGCTGGCGACCGCGGCTCGACTCGCGATGACGTTGT
TGCAGTGCGTGTTGCCGCGCCAGGTTTTGATACAGCTCCTCCTGAGTTGTGCTCATGGTTGGAC
```

- We know which variety each tag comes from because it has a barcode.
- Barcode is trimmed after identification.

Tag-by-Taxa Grid:

All tags that appear more than (10) times:
(typically, millions of tags)

All varieties
(Taxa)

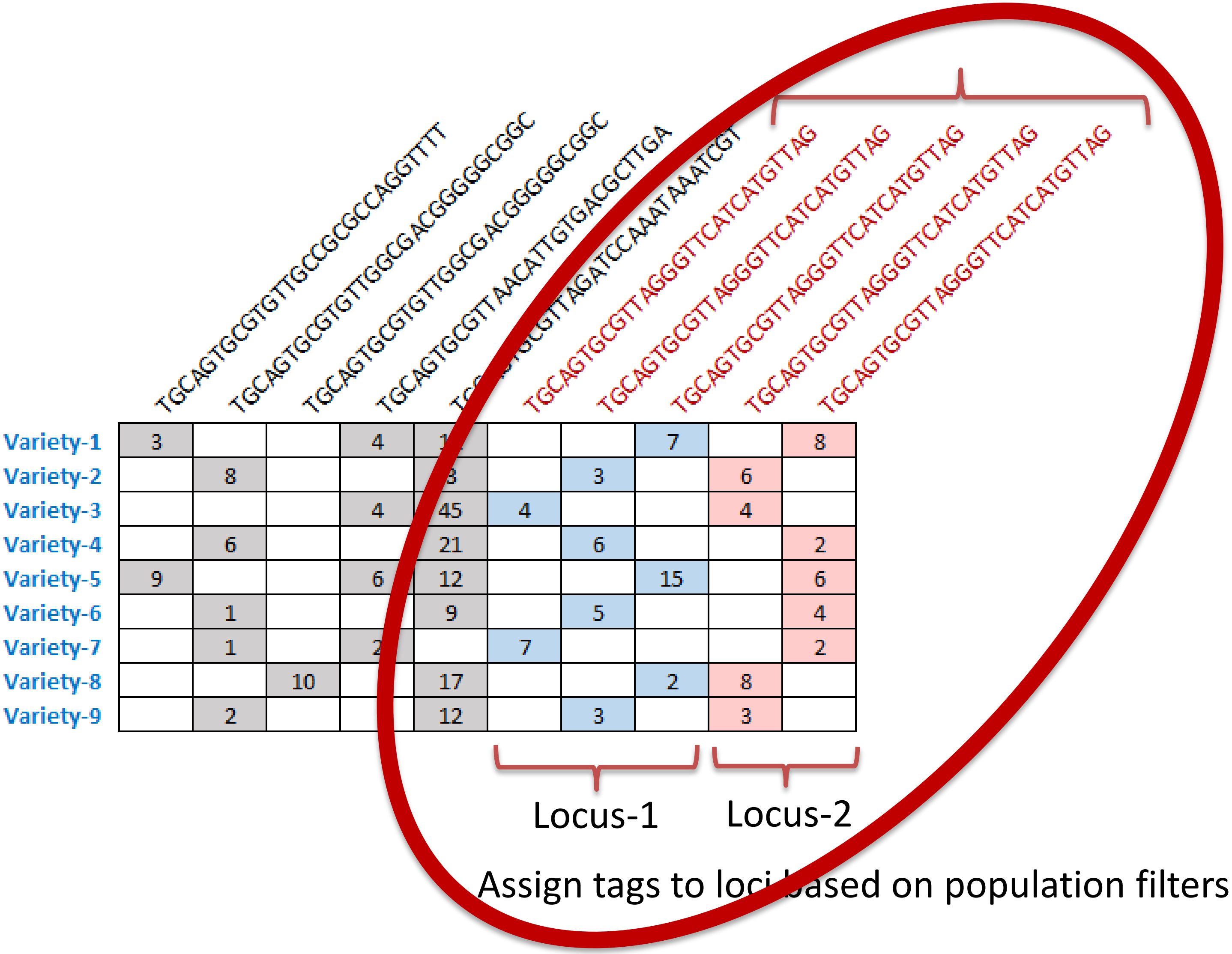
	TGCAGTGCGTGTTGCCGCGCCAGGTTTT	TGCAGTGCGTGTTGGCGACGGGGGCGGC	TGCAGTGCGTGTTGGCGACGGGGGCGGC	TGCAGTGCGTTAACATTGTGACGCTTGA	TGCAGTGCGTTAGATCCAAATAAATCGT	TGCAGTGCGTTAGGGTTCATCATGTTAG	TGCAGTGCGTTAGGGTTCATCATGTTAG	TGCAGTGCGTTAGGGTTCATCATGTTAG	TGCAGTGCGTTAGGGTTCATCATGTTAG	TGCAGTGCGTTAGGGTTCATCATGTTAG
Variety-1	3			4	12			7		8
Variety-2		8			3		3		6	
Variety-3				4	45	4			4	
Variety-4		6			21		6			2
Variety-5	9			6	12			15		6
Variety-6		1			9		5			4
Variety-7		1		2		7				2
Variety-8			10		17			2	8	
Variety-9		2			12		3		3	

- Count the number of times we see a given tag in each variety
- We CAN use this directly as a presence-absence genotype grid
- Better if we can assign tags to discrete diploid loci ……

What Haplotag does....

- These tags belong to a “cluster” Based on sequence similarity.

* They may be an A-C-D series.



	TGCAGTGCGTGTTGCCGCGCCAGGTTTT	TGCAGTGCGTGTTGGCGACGGGGGCGGC	TGCAGTGCGTGTTGGCGACGGGGGCGGC	TGCAGTGCGTTAACATTGTGACGCTTGA	TG[illegible]CGTTAGATCCAAATAAAATCGT	TGCAGTGCGTTAGGGTTCATCATGTTAG	TGCAGTGCGTTAGGGTTCATCATGTTAG	TGCAGTGCGTTAGGGTTCATCATGTTAG	TGCAGTGCGTTAGGGTTCATCATGTTAG	TGCAGTGCGTTAGGGTTCATCATGTTAG
Variety-1	3			4	12			7		8
Variety-2		8			3		3		6	
Variety-3				4	45	4			4	
Variety-4		6			21		6			2
Variety-5	9			6	12			15		6
Variety-6		1			9		5			4
Variety-7		1		2		7				2
Variety-8			10		17			2	8	
Variety-9		2			12		3		3	

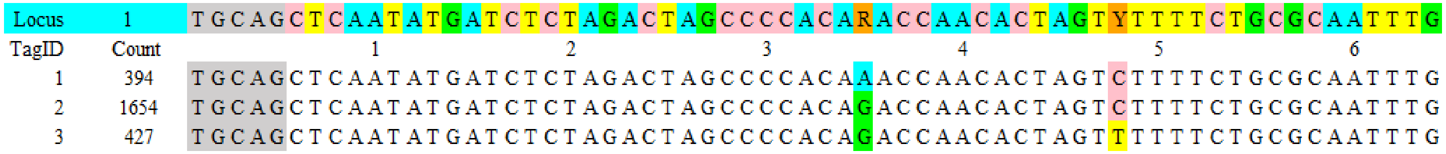
Assign tags to loci based on population filters

Haplotag passport

… in your browser: Haplotag.aowc.ca/example

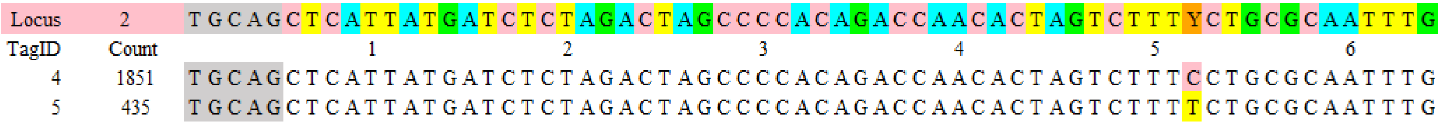
and on T3/Oat public database

1

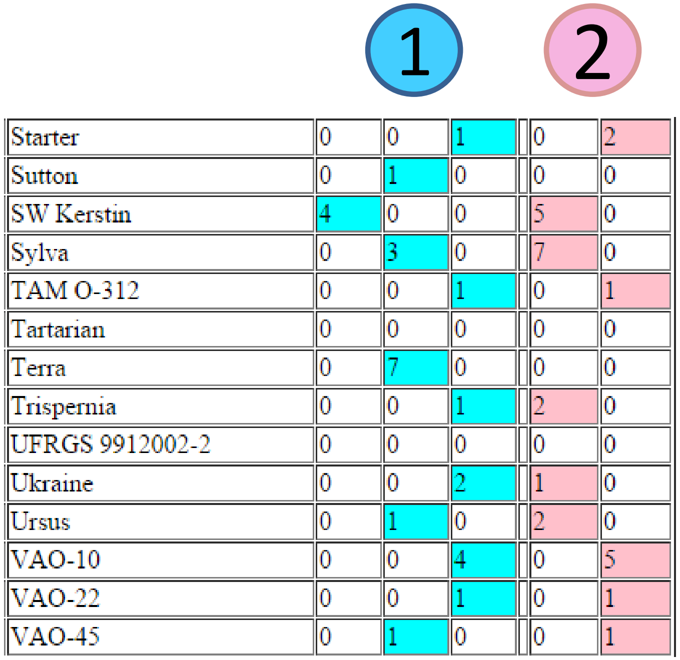


Locus	1	TGCAGCTCAATATGATCTCTAGACTAGCCCCACARACCAACACTAGTYTTTTCTGCGCAATTTG
TagID	Count	1 2 3 4 5 6
1	394	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAAACCAACACTAGTCTTTTCTGCGCAATTTG
2	1654	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTTCTGCGCAATTTG
3	427	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTTTTTTCTGCGCAATTTG

2

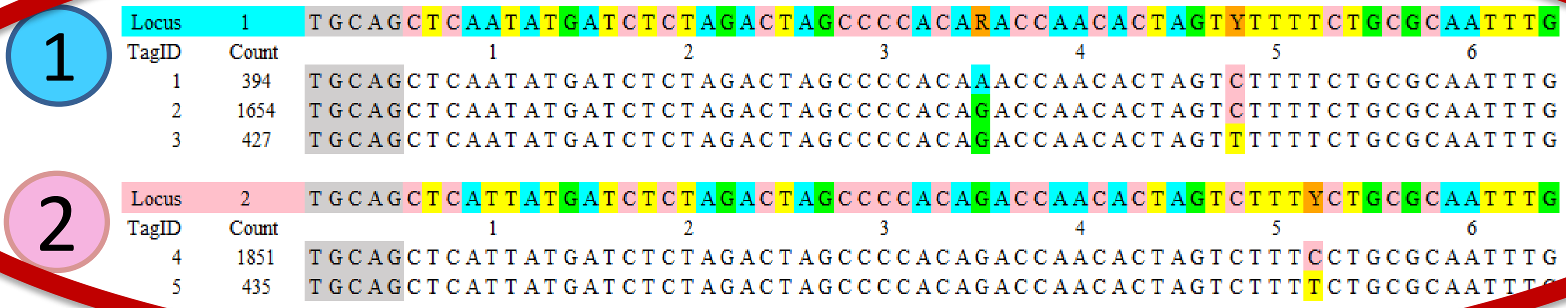


Locus	2	TGCAGCTCATTATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTYCTGCGCAATTTG
TagID	Count	1 2 3 4 5 6
4	1851	TGCAGCTCATTATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTCCTGCGCAATTTG
5	435	TGCAGCTCATTATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTTCTGCGCAATTTG



	1			2	
Starter	0	0	1	0	2
Sutton	0	1	0	0	0
SW Kerstin	4	0	0	5	0
Sylva	0	3	0	7	0
TAM O-312	0	0	1	0	1
Tartarian	0	0	0	0	0
Terra	0	7	0	0	0
Trispernia	0	0	1	2	0
UFRGS 9912002-2	0	0	0	0	0
Ukraine	0	0	2	1	0
Ursus	0	1	0	2	0
VAO-10	0	0	4	0	5
VAO-22	0	0	1	0	1
VAO-45	0	1	0	0	1

Haplotag.aowc.ca/example



①

Locus	1	TGCAGCTCAATATGATCTCTAGACTAGCCCCACARACCAACACTAGTYTTTTCTGCGCAATTTG
TagID	Count	1 2 3 4 5 6
1	394	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAAACCAACACTAGTCTTTTCTGCGCAATTTG
2	1654	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTTCTGCGCAATTTG
3	427	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTTTTTTCTGCGCAATTTG

②

Locus	2	TGCAGCTCATTATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTYCTGCGCAATTTG
TagID	Count	1 2 3 4 5 6
4	1851	TGCAGCTCATTATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTCCTGCGCAATTTG
5	435	TGCAGCTCATTATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTTCTGCGCAATTTG

Cluster

With 2 loci

Haplotag.aowc.ca/example

1

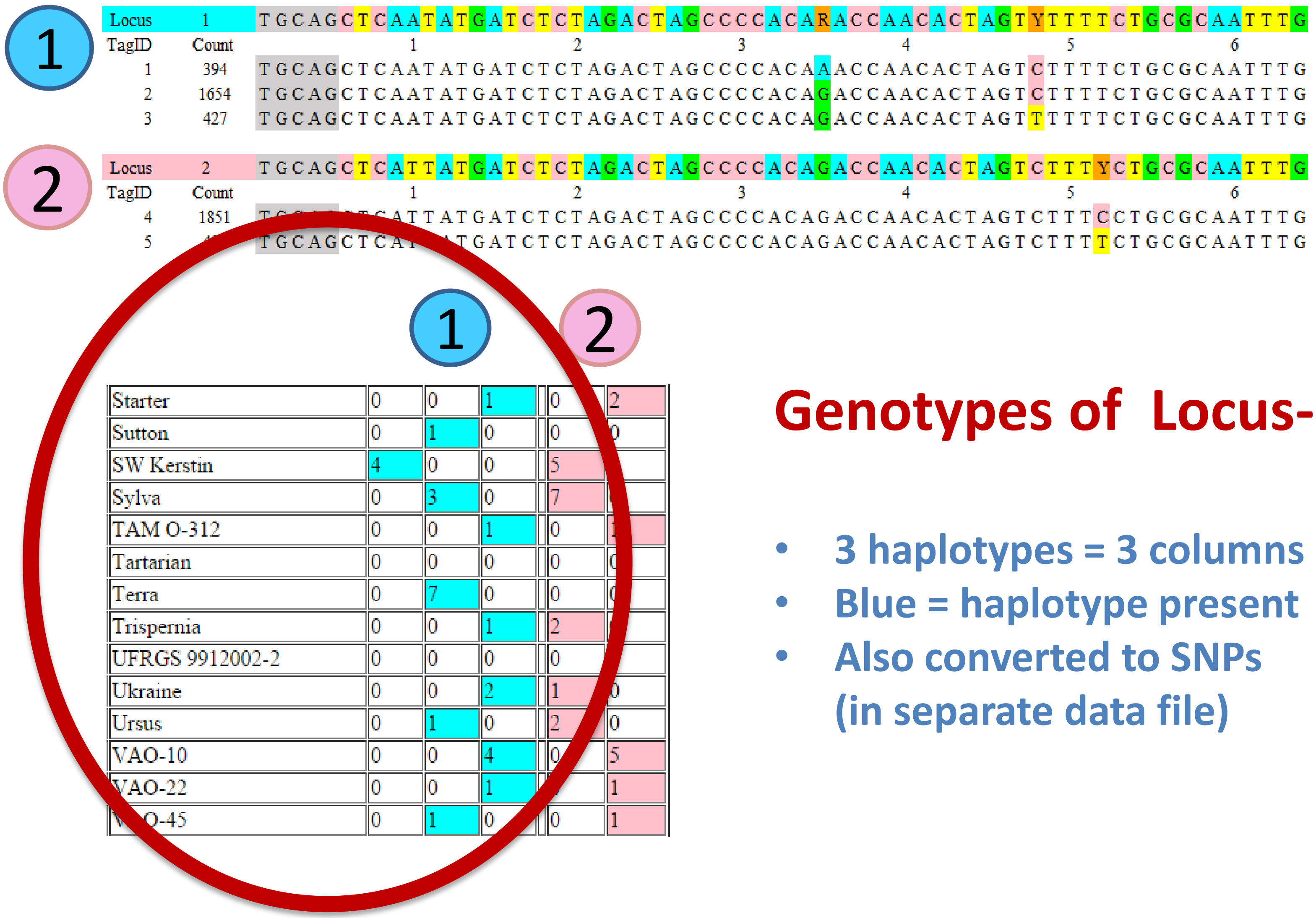
Locus	1	TGCAGCTCAATATGATCTCTAGACTAGCCCCACARACCAACACTAGTYTTTTCTGCGCAATTTG
TagID	Count	1 2 3 4 5 6
1	394	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAAACCAACACTAGTCTTTTCTGCGCAATTTG
2	1654	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTTCTGCGCAATTTG
3	427	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTTTTTTCTGCGCAATTTG

Locus-1

- 3 tag-level haplotypes
- 2 SNPs

Haplotag.aowc.ca/example

Haplotag passport



1

Locus	1	TGCAGCTCAATATGATCTCTAGACTAGCCCCACARACCAACACTAGTYTTTTCTGCGCAATTTG
TagID	Count	1 2 3 4 5 6
1	394	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAAACCAACACTAGTCTTTTCTGCGCAATTTG
2	1654	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTTCTGCGCAATTTG
3	427	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTTTTTTCTGCGCAATTTG

2

Locus	2	TGCAGCTCATTATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTYCTGCGCAATTTG
TagID	Count	1 2 3 4 5 6
4	1851	[illegible]CATTATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTCCTGCGCAATTTG
5	4[illegible]	TGCAGCTCA[illegible]TGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTTCTGCGCAATTTG

		1			2
Starter	0	0	1	0	2
Sutton	0	1	0	0	0
SW Kerstin	4	0	0	5	[illegible]
Sylva	0	3	0	7	0
TAM O-312	0	0	1	0	1
Tartarian	0	0	0	0	0
Terra	0	7	0	0	0
Trispernia	0	0	1	2	0
UFRGS 9912002-2	0	0	0	0	[illegible]
Ukraine	0	0	2	1	0
Ursus	0	1	0	2	0
VAO-10	0	0	4	0	5
VAO-22	0	0	1	[illegible]	1
V[illegible]O-45	0	1	0	0	1

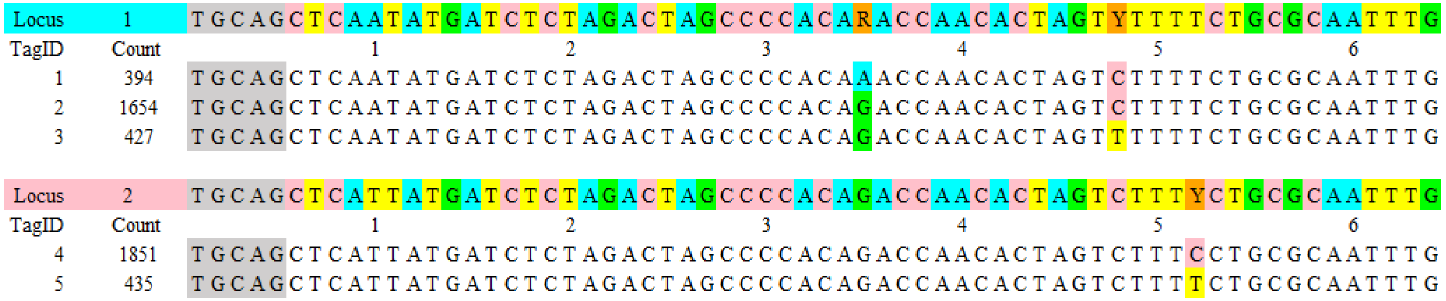
Genotypes of Locus-1

- 3 haplotypes = 3 columns
- Blue = haplotype present
- Also converted to SNPs
  (in separate data file)

Haplotag passport

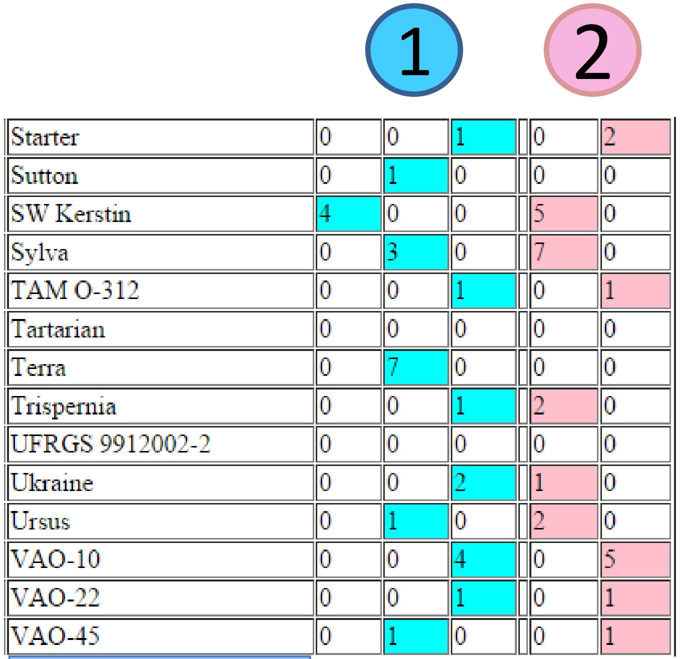
1

2



```
Locus   1       TGCAGCTCAATATGATCTCTAGACTAGCCCCACARACCAACACTAGTYTTTTCTGCGCAATTTG
TagID   Count             1         2         3         4         5         6
    1     394   TGCAGCTCAATATGATCTCTAGACTAGCCCCACAAACCAACACTAGTCTTTTCTGCGCAATTTG
    2    1654   TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTTCTGCGCAATTTG
    3     427   TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTTTTTTCTGCGCAATTTG

Locus   2       TGCAGCTCATTATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTYCTGCGCAATTTG
TagID   Count             1         2         3         4         5         6
    4    1851   TGCAGCTCATTATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTCCTGCGCAATTTG
    5     435   TGCAGCTCATTATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTTCTGCGCAATTTG
```



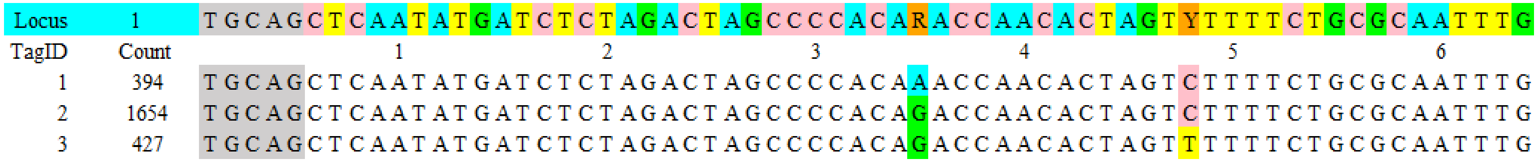
	1			2	
Starter	0	0	1	0	2
Sutton	0	1	0	0	0
SW Kerstin	4	0	0	5	0
Sylva	0	3	0	7	0
TAM O-312	0	0	1	0	1
Tartarian	0	0	0	0	0
Terra	0	7	0	0	0
Trispernia	0	0	1	2	0
UFRGS 9912002-2	0	0	0	0	0
Ukraine	0	0	2	1	0
Ursus	0	1	0	2	0
VAO-10	0	0	4	0	5
VAO-22	0	0	1	0	1
VAO-45	0	1	0	0	1

4657

Diversity

- 4657 varieties in cultivated oat
  - CORE project
  - Public Oat Genotyping
  - Ottawa breeding program
- Data on……
  - 156,000 clusters/passports
  - 165,000 loci
  - 353,000 haplotypes
  - 241,000 SNPs

Haplotag passport – 12 CORE mapping populations



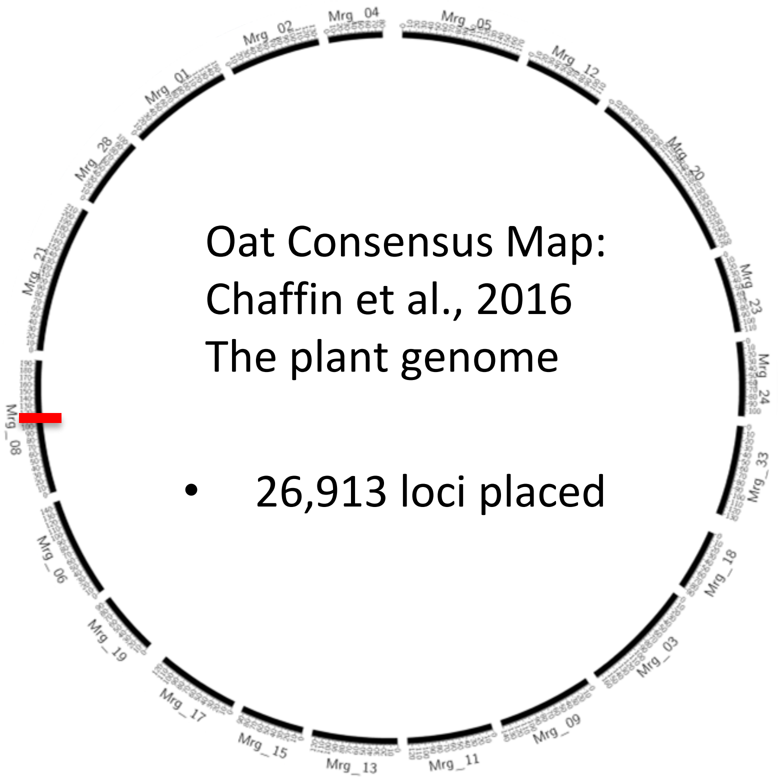
Locus	1	TGCAGCTCAATATGATCTCTAGACTAGCCCCACARACCAACACTAGTYTTTTCTGCGCAATTTG
TagID	Count	1 2 3 4 5 6
1	394	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAAACCAACACTAGTCTTTTCTGCGCAATTTG
2	1654	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTTCTGCGCAATTTG
3	427	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTTTTTTCTGCGCAATTTG

Haplotypes 2 and 3 segregate in KxO

KO.023	0	1	0
KO.024	0	5	0
KO.026	0	0	1
KO.027	0	4	0
KO.028	0	2	0
KO.030	0	0	4
KO.031	0	1	0
KO.032	0	3	0
KO.033	0	0	5
KO.035	0	2	0

Haplotypes 1 and 2 segregate in HxZ

HZ.003	0	4	0
HZ.004	1	0	0
HZ.005	0	1	0
HZ.006	5	0	0
HZ.007	3	0	0
HZ.008	0	3	0
HZ.009	0	3	0
HZ.010	0	0	0
HZ.011	2	0	0
HZ.012	0	0	0
HZ.013	0	6	0
HZ.014	0	1	0



Oat Consensus Map:
Chaffin et al., 2016
The plant genome

- 26,913 loci placed

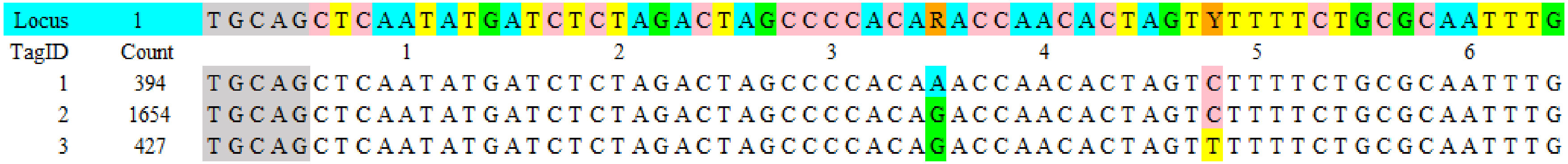
Wake up!

The rest is

- Short
- fun (?)
- Unpublished !

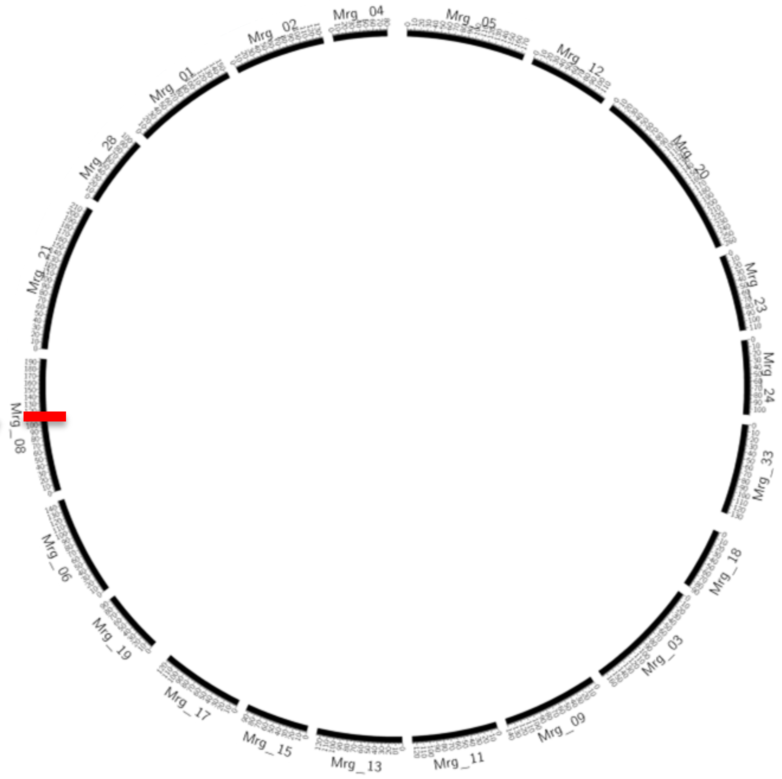
- Everything I just showed is in <u>cultivated hexaploid oat</u>
- What about doing GBS in other *Avena* species ??

Can we find these “cultivated haplotypes” in other oat species ?



Locus	1	TGCAGCTCAATATGATCTCTAGACTAGCCCCACARACCAACACTAGTYTTTTCTGCGCAATTTG
TagID	Count	1 2 3 4 5 6
1	394	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAAACCAACACTAGTCTTTTCTGCGCAATTTG
2	1654	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTCTTTTCTGCGCAATTTG
3	427	TGCAGCTCAATATGATCTCTAGACTAGCCCCACAGACCAACACTAGTTTTTTCTGCGCAATTTG

1. Can the number of matched haplotypes (across all loci) tell us which species are closest to cultivated oat ?

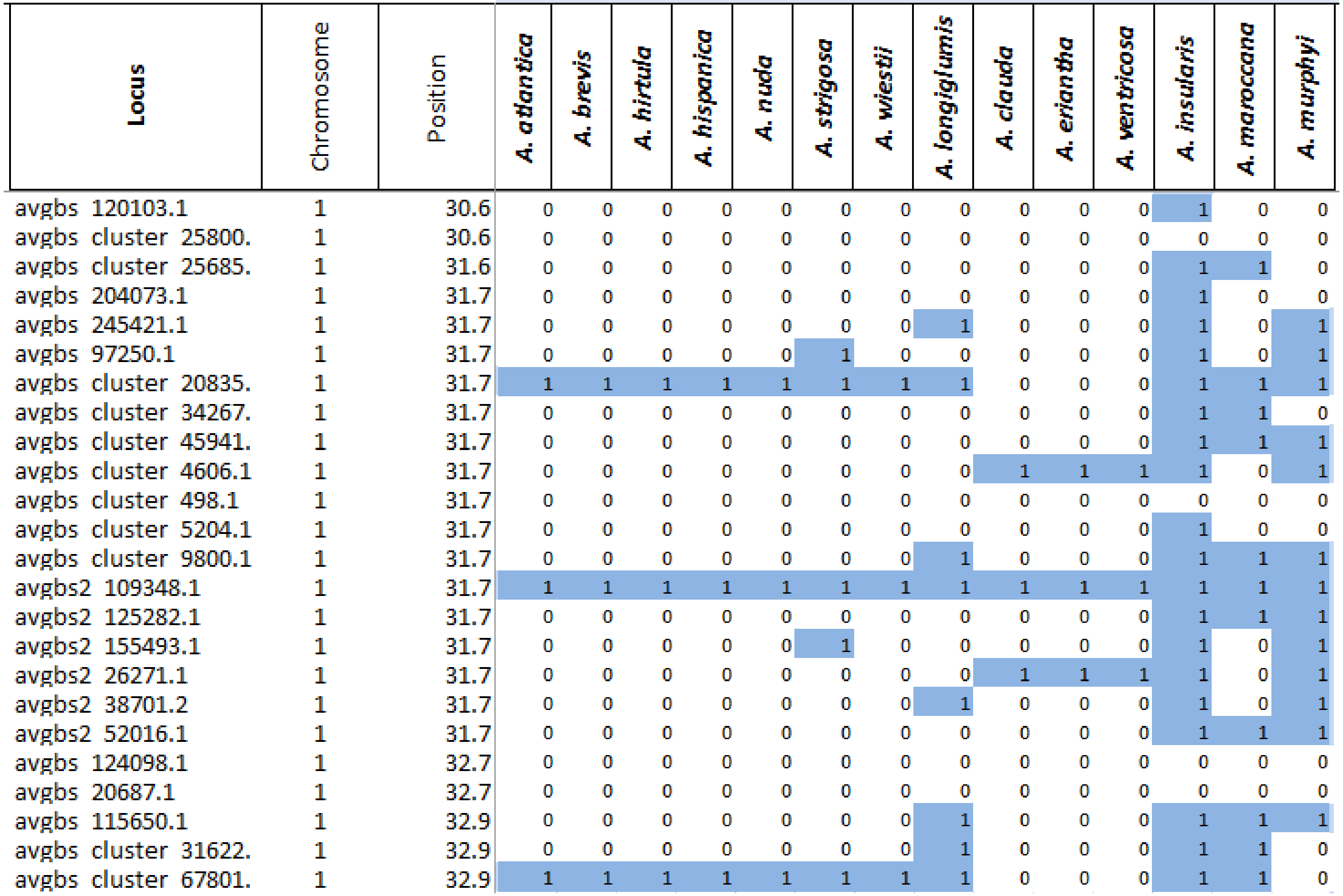


2. Can we infer the origin of the hexaploid chromosome that contains this locus ?

Matching cultivated haplotypes in wild oat species

26,913 loci

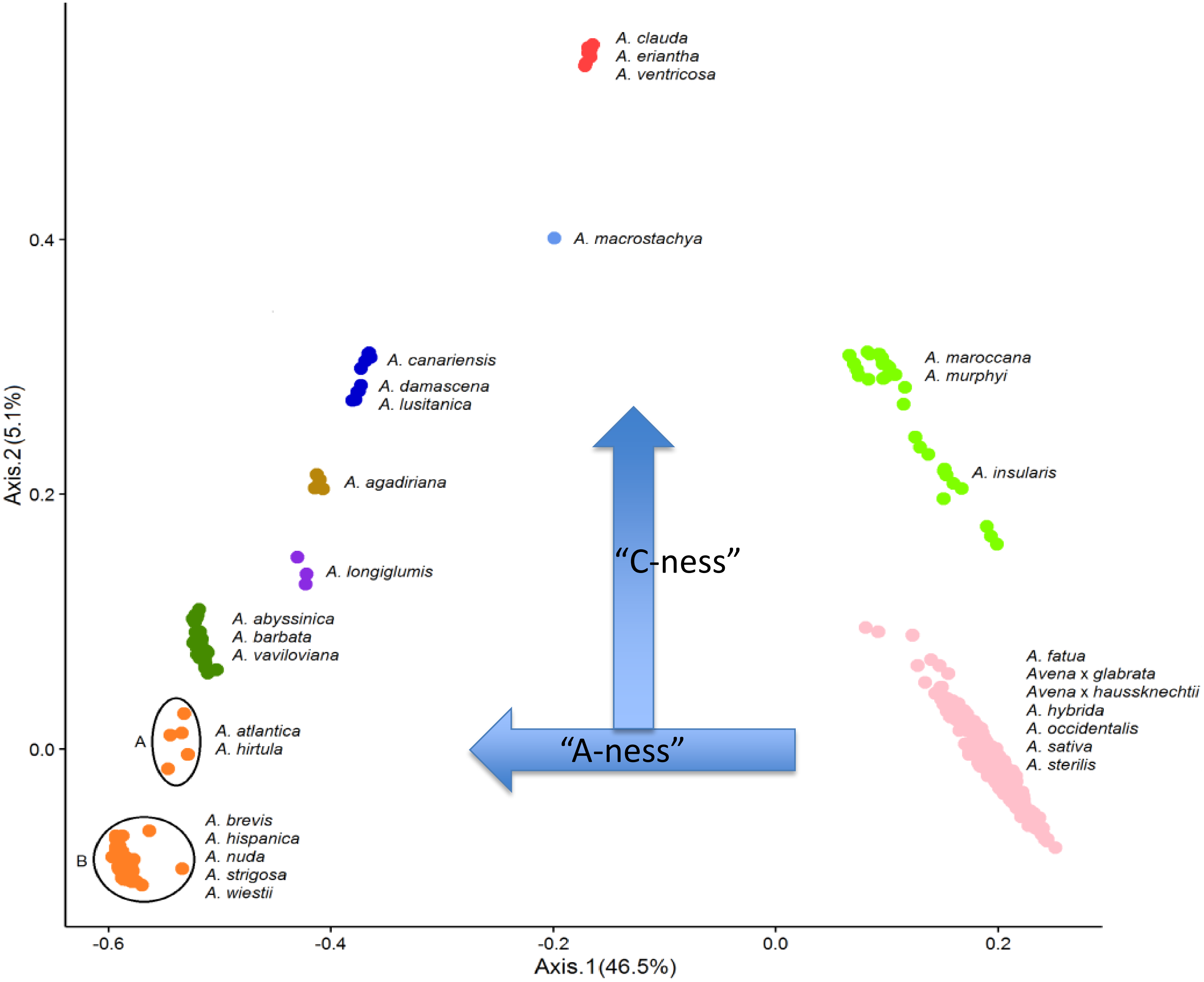
27 oat species, 266 accessions



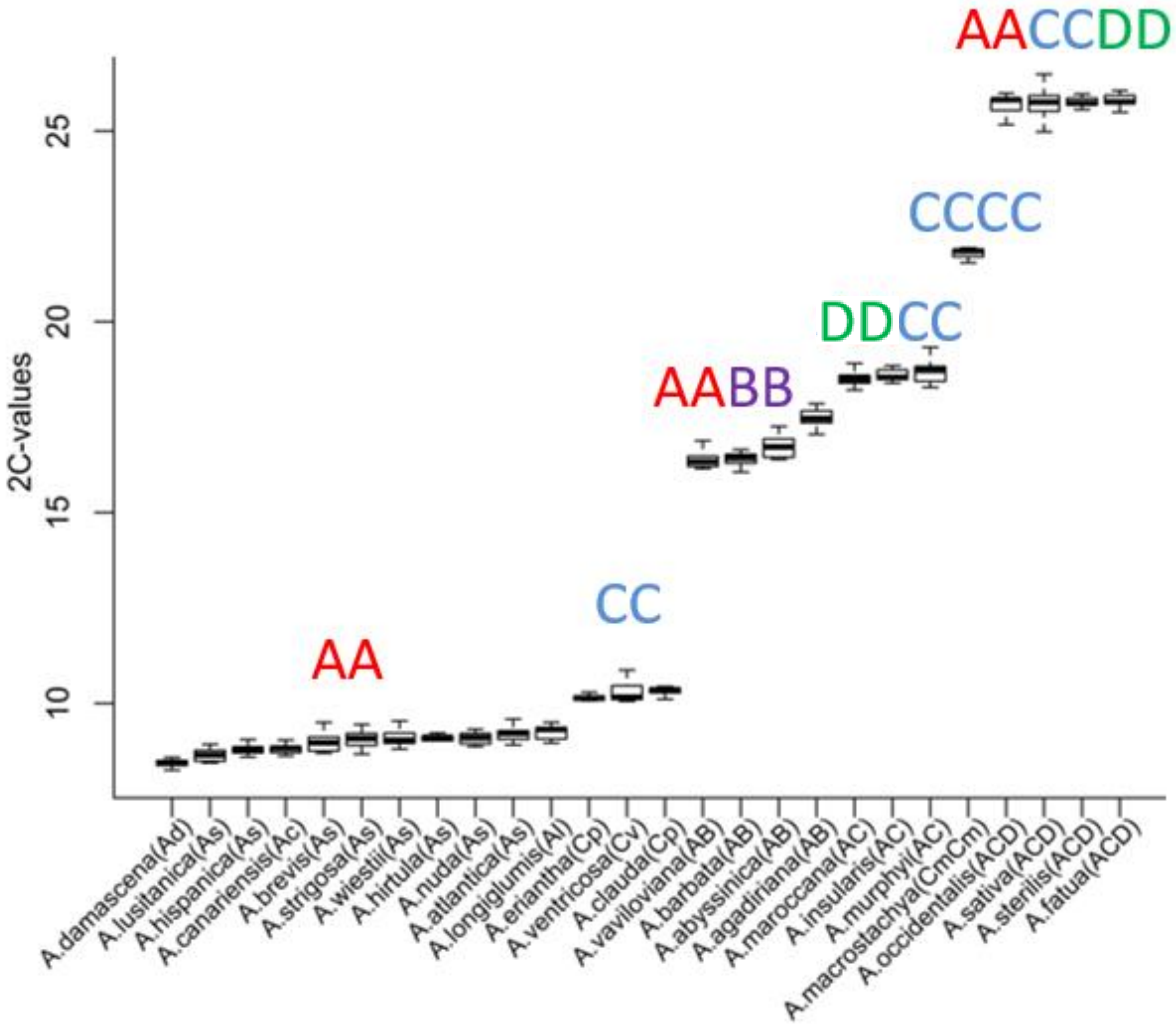
Locus	Chromosome	Position	*A. atlantica*	*A. brevis*	*A. hirtula*	*A. hispanica*	*A. nuda*	*A. strigosa*	*A. wiestii*	*A. longiglumis*	*A. clauda*	*A. eriantha*	*A. ventricosa*	*A. insularis*	*A. maroccana*	*A. murphyi*
avgbs 120103.1	1	30.6	0	0	0	0	0	0	0	0	0	0	0	1	0	0
avgbs cluster 25800.	1	30.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
avgbs cluster 25685.	1	31.6	0	0	0	0	0	0	0	0	0	0	0	1	1	0
avgbs 204073.1	1	31.7	0	0	0	0	0	0	0	0	0	0	0	1	0	0
avgbs 245421.1	1	31.7	0	0	0	0	0	0	0	1	0	0	0	1	0	1
avgbs 97250.1	1	31.7	0	0	0	0	0	1	0	0	0	0	0	1	0	1
avgbs cluster 20835.	1	31.7	1	1	1	1	1	1	1	1	0	0	0	1	1	1
avgbs cluster 34267.	1	31.7	0	0	0	0	0	0	0	0	0	0	0	1	1	0
avgbs cluster 45941.	1	31.7	0	0	0	0	0	0	0	0	0	0	0	1	1	1
avgbs cluster 4606.1	1	31.7	0	0	0	0	0	0	0	0	1	1	1	1	0	1
avgbs cluster 498.1	1	31.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
avgbs cluster 5204.1	1	31.7	0	0	0	0	0	0	0	0	0	0	0	1	0	0
avgbs cluster 9800.1	1	31.7	0	0	0	0	0	0	0	1	0	0	0	1	1	1
avgbs2 109348.1	1	31.7	1	1	1	1	1	1	1	1	1	1	1	1	1	1
avgbs2 125282.1	1	31.7	0	0	0	0	0	0	0	0	0	0	0	1	1	1
avgbs2 155493.1	1	31.7	0	0	0	0	0	1	0	0	0	0	0	1	0	1
avgbs2 26271.1	1	31.7	0	0	0	0	0	0	0	0	1	1	1	1	0	1
avgbs2 38701.2	1	31.7	0	0	0	0	0	0	0	1	0	0	0	1	0	1
avgbs2 52016.1	1	31.7	0	0	0	0	0	0	0	0	0	0	0	1	1	1
avgbs 124098.1	1	32.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
avgbs 20687.1	1	32.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
avgbs 115650.1	1	32.9	0	0	0	0	0	0	0	1	0	0	0	1	1	1
avgbs cluster 31622.	1	32.9	0	0	0	0	0	0	0	1	0	0	0	1	1	0
avgbs cluster 67801.	1	32.9	1	1	1	1	1	1	1	1	0	0	0	1	1	0

- Convert to 1/0 grid
- PCoA analysis of accessions

PCoA analysis: 266 accessions from 27 species of Avena

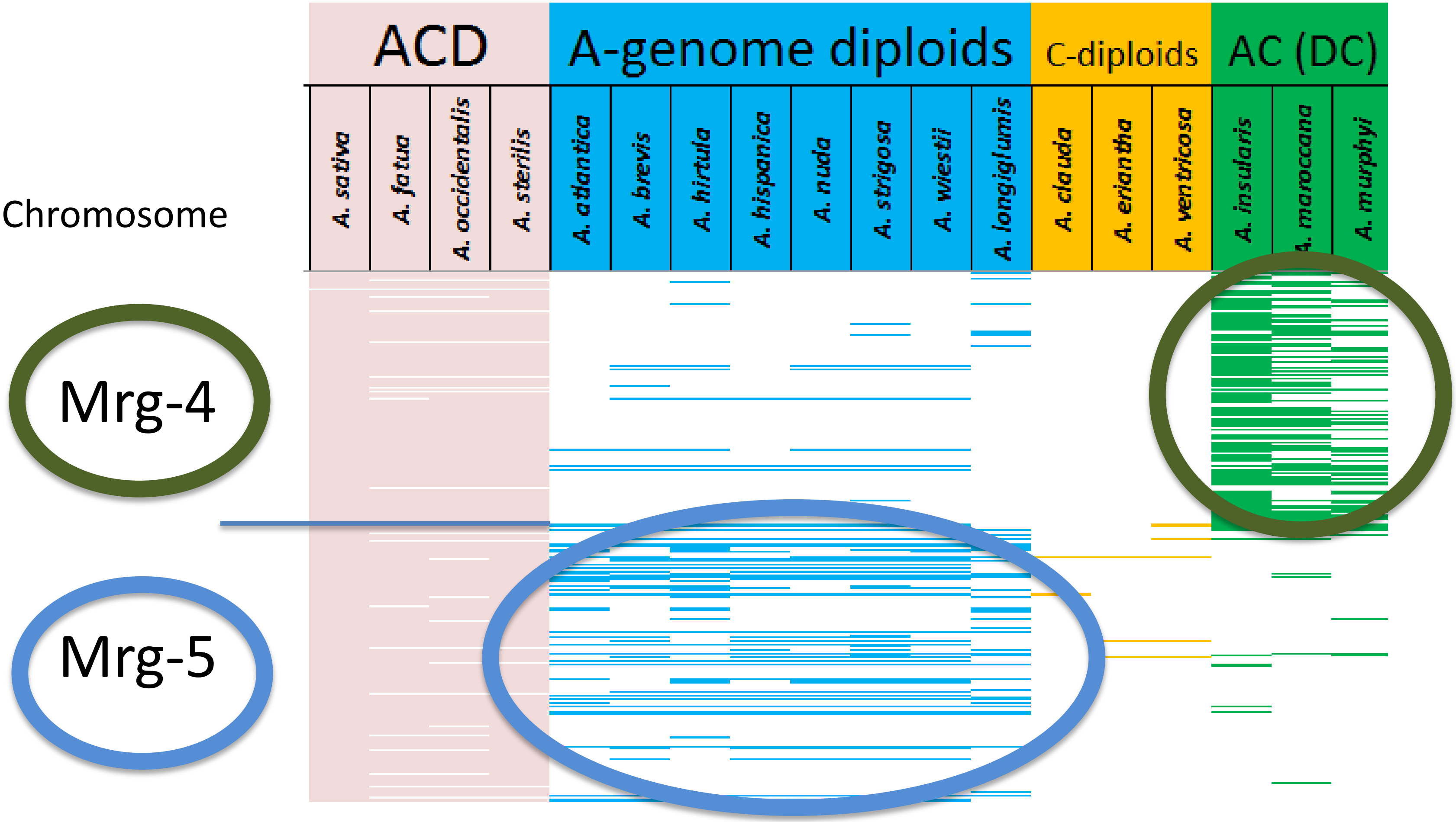


Genome size of 26 Avena species

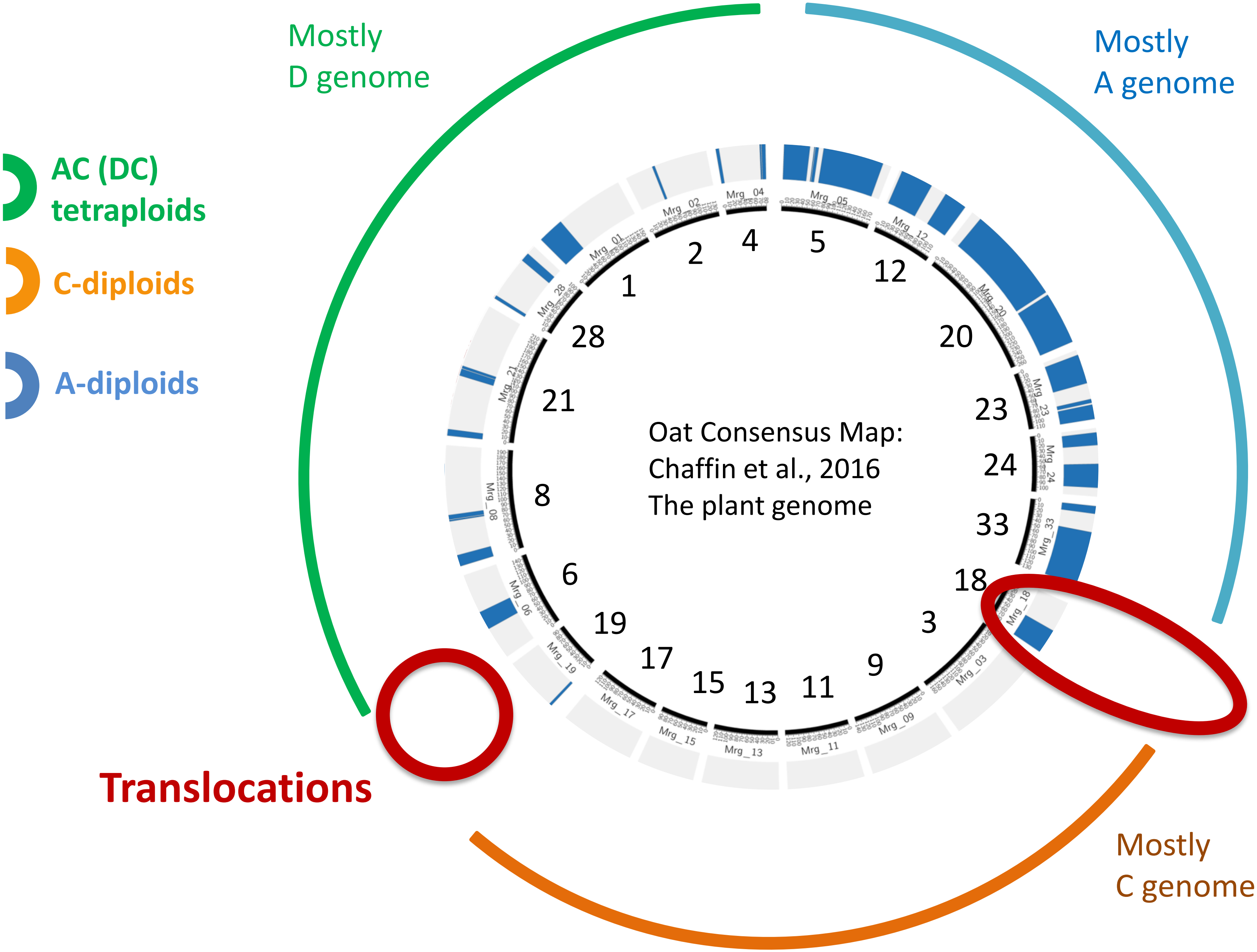


Yan, et al. and Tinker (2016) Genome 59

Cultivated haplotypes - matching by ancestral groups



- Combine by ancestral groups
- Smooth this over 30cM Window, apply threshold



Mostly
D genome

Mostly
A genome

AC (DC)
tetraploids

C-diploids

A-diploids

Oat Consensus Map:
Chaffin et al., 2016
The plant genome

Translocations

Mostly
C genome

Summary

Haplotag applied in cultivated hexaploid on a large scale

- Improved consensus map, diversity analysis, GWAS, and GS

Haplotypes from cultivated oat in 27 species - What's new?

- First <u>genetic</u> distinction among three genomes of cultivated oat
- Assignments match perfectly with the 9 previous assigned chromosomes
- Allows completion/validation of chromosome assignments
- Demonstrates that the tetraploids contributed the D genome
- Identifies translocations on the map
- Avena insularis is the closest non-hexaploid species to cultivated oat

- Will assist with hexaploid genome mapping and sequencing
- Implications for germplasm conservation and allele mining

Thank You!

Collaborators

Yuanying Peng
Tim Langdon
Axel Diederichsen
Catherine Howarth
Bob Latta
Yong-Bi Fu
Rick Jellen
Kathy Klos
Jessica Schlueter
… and many others

North American Millers' Association
NAMA

Prairie Oat Growers Association

USDA

Agriculture and Agri-Food Canada

GENERAL MILLS

PEPSICO

Tinker lab:

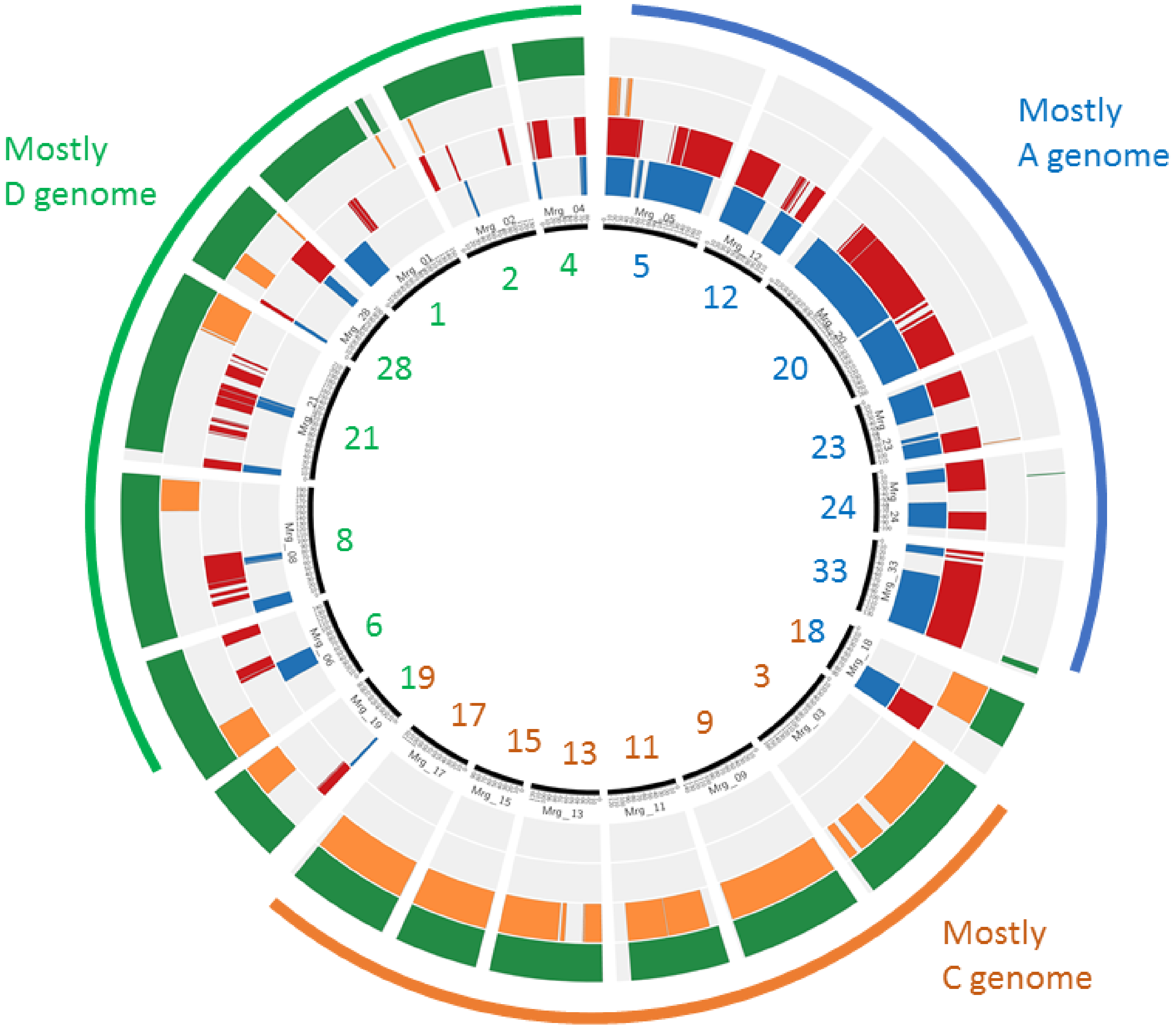
Charlene Wight
Wubishet Bekele
Honghai Yan
Phil Couroux
Jiro Hattori
Pingping Zhou
Biniam Hizbai

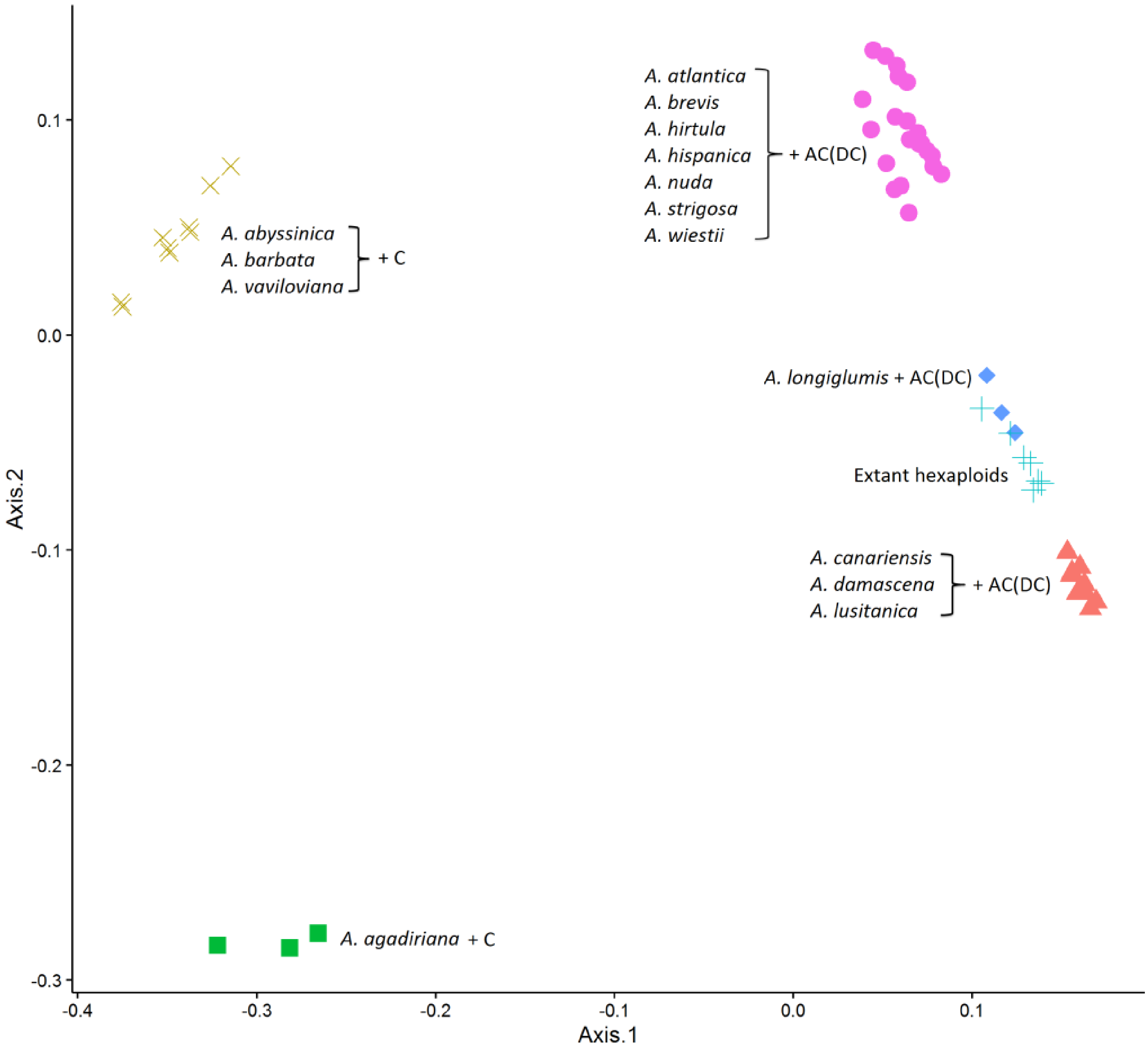
If you miss my talk tomorrow (concurrent session):

	Oat Global	Oatglobal.org
Poster	T3/Oat database	triticeaetoolbox.org/oat/
	Public Oat Genotyping	
	GrainGenes	http://wheat.pw.usda.gov/GG3/
Poster Poster	Oat Newsletter + Twitter	Oatnews.org
	Oat Rust	z.umn.edu/oatrustinitiative

Table 2. Previous and current physical assignment of 21 chromosome representations from the hexaploid oat consensus map.

Consensus chromosome[a]	Confirmed assignment[a]	Previous assignment[b]	Diploid assignment[c]	New genome assignment [d]
Mrg05	16A	16A, 1C	AD	A
Mrg12		13A	AD	A
Mrg20		19A	AD	A
Mrg23	11A	11A	AD	A
Mrg24		8A (14D)	AD	A
Mrg33		15A	AD	A
Mrg18		7C-17A	C	C / A
Mrg03	4C	4C (10D)	C	C
Mrg09	6C	6C	C	C
Mrg11		1C	C / A	C
Mrg13		(20D)	C	C
Mrg15	2C	2C (10D)	C	C
Mrg17		3C	C	C
Mrg19	21D	21D	C / AD	C / D
Mrg06		14D	AD / C	D / C
Mrg08	12D	12D	AD / C	D / C
Mrg21		(16A)	AD / C	D / C
Mrg28		7C-(17A)	C	D / C
Mrg01		(5C)	AD	D
Mrg02	9D	9D	AD	D
Mrg04	18D	18D	AD	D


Mostly D genome
Mostly A genome
Mostly C genome
Mrg_04
4
Mrg_05
5
Mrg_12
12
Mrg_20
20
Mrg_23
23
Mrg_24
24
Mrg_33
33
Mrg_18
18
Mrg_03
3
Mrg_09
9
Mrg_11
11
Mrg_13
13
Mrg_15
15
Mrg_17
17
Mrg_19
19
Mrg_06
6
Mrg_08
8
Mrg_21
21
Mrg_28
28
Mrg_01
1
Mrg_02
2


A. atlantica
A. brevis
A. hirtula
A. hispanica
A. nuda
A. strigosa
A. wiestii
+ AC(DC)
A. abyssinica
A. barbata
A. vaviloviana
+ C
A. longiglumis + AC(DC)
Extant hexaploids
A. canariensis
A. damascena
A. lusitanica
+ AC(DC)
A. agadiriana + C
0.1
0.0
-0.1
-0.2
-0.3
Axis.2
-0.4
-0.3
-0.2
-0.1
0.0
0.1
Axis.1

Haplotag Summary

Tinker et al. (2016) Genes, Genomes and Genetics

Features

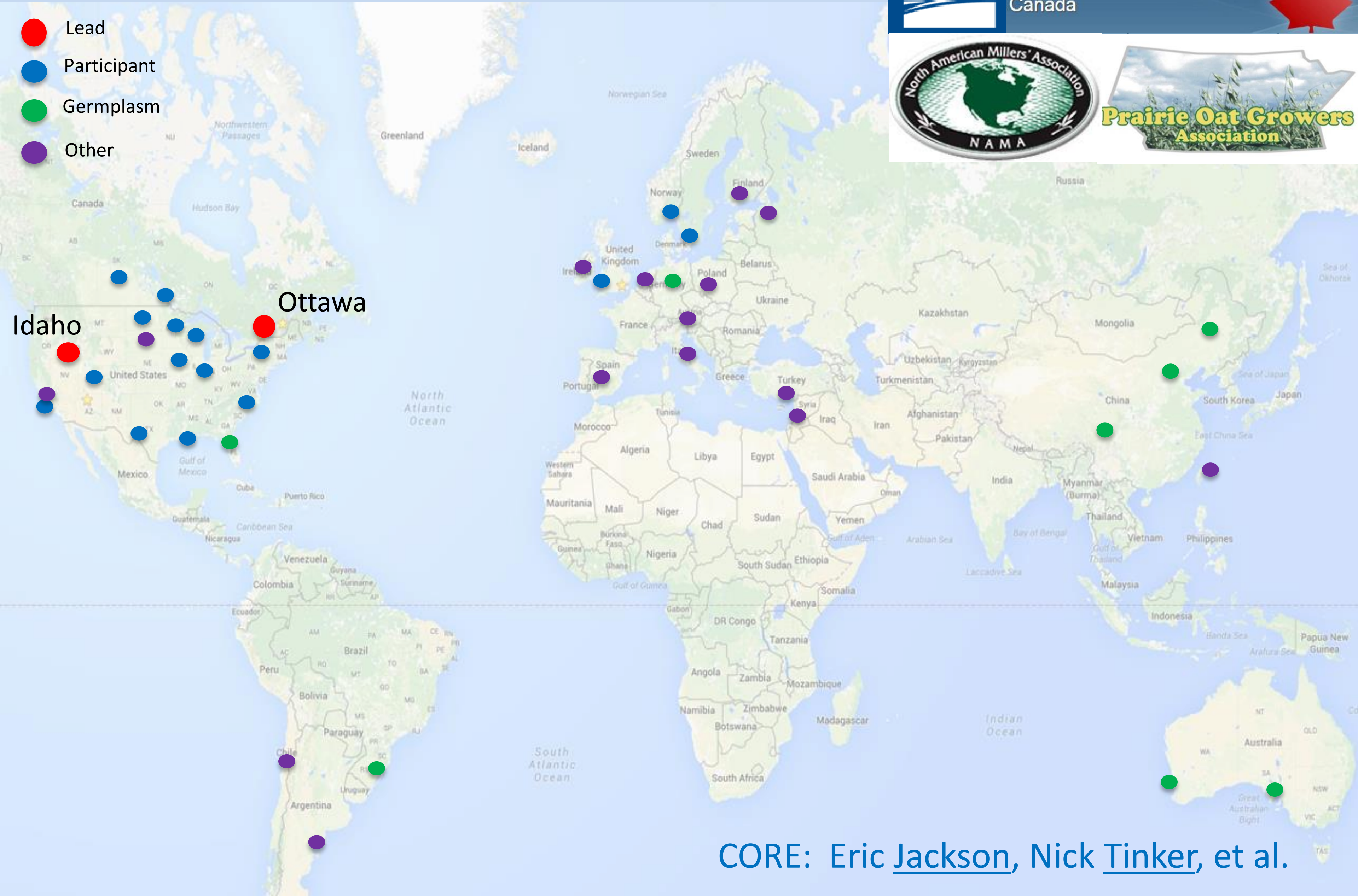
- Runs on Windows, Free!
- Works in polyploid without a reference genome
- Graphical passport for each cluster of loci
- Saves genotype data as SNPs (bi-allelic) or haplotypes (multi-allelic)
- Two modes:
  - Cluster and locus discovery (slow, for new species)
  - Production mode (fast – looks for tags from pre-defined loci)

Collaborative Oat Research Enterprise (CORE) 2009-2014

USDA
Agriculture and Agri-Food Canada

North American Millers' Association
NAMA

Prairie Oat Growers Association



CORE: Eric Jackson, Nick Tinker, et al.