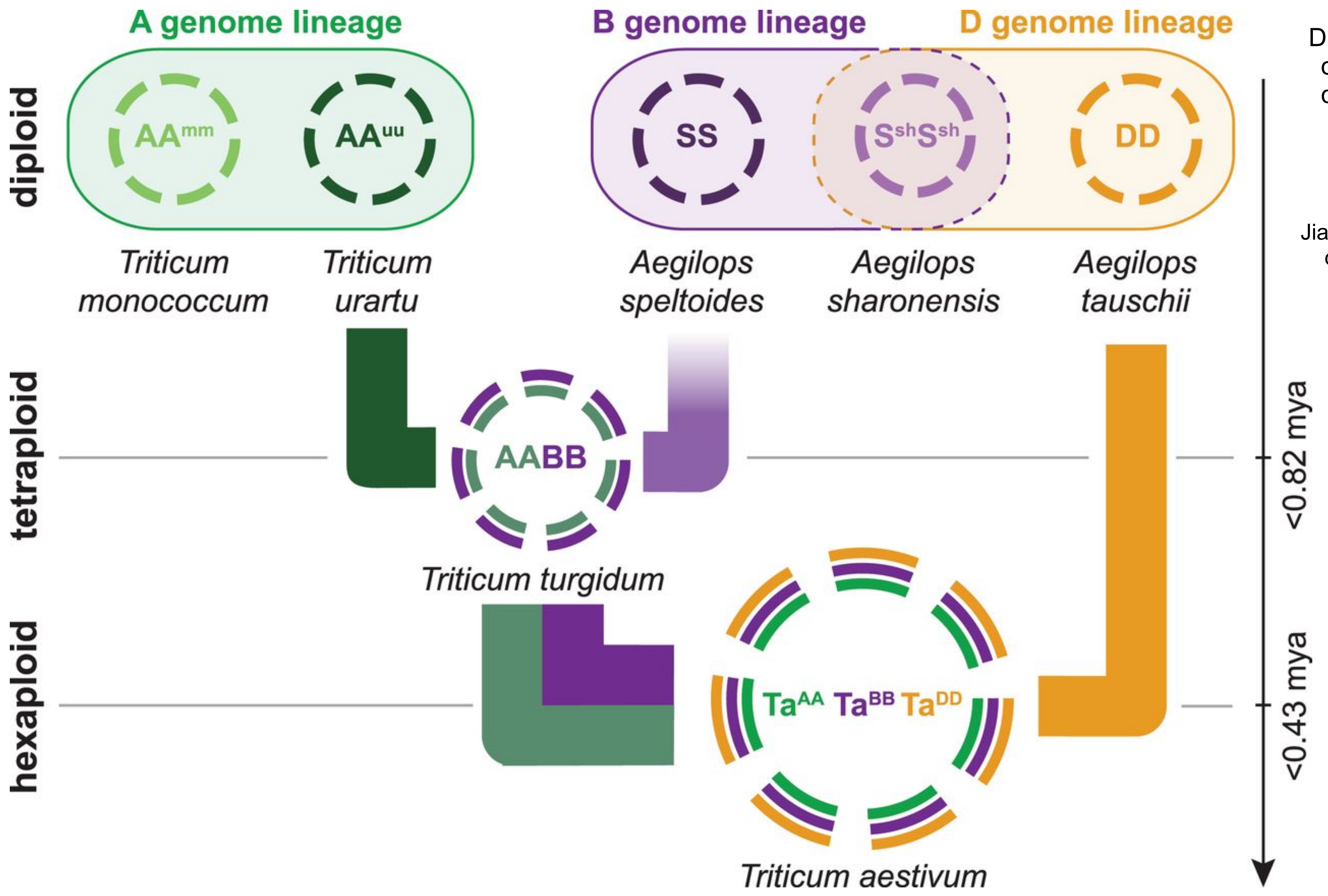
Development of Genetic and Genomic Resources to Evaluate Wheat Organellar Genome Variants and Their Functional Implications

Katie L. Liberatore[1,2], Marisa E. Miller[1,2], Kevin A. T. Silverstein[3], Castle Raley[4], Bao Tran[4] and Shahryar F. Kianian[1,2]
(1)Cereal Disease Laboratory, United States Department of Agriculture-Agricultural Research Service, St. Paul, MN,
(2)Department of Plant Pathology, University of Minnesota, St. Paul, MN,
(3)Supercomputing Institute, University of Minnesota, Minneapolis, MN,
(4)Leidos Biomedical Research, Inc; Frederick National Laboratory for Cancer Research, Frederick, MD

Nuclear genome diversity was reduced during evolution and domestication



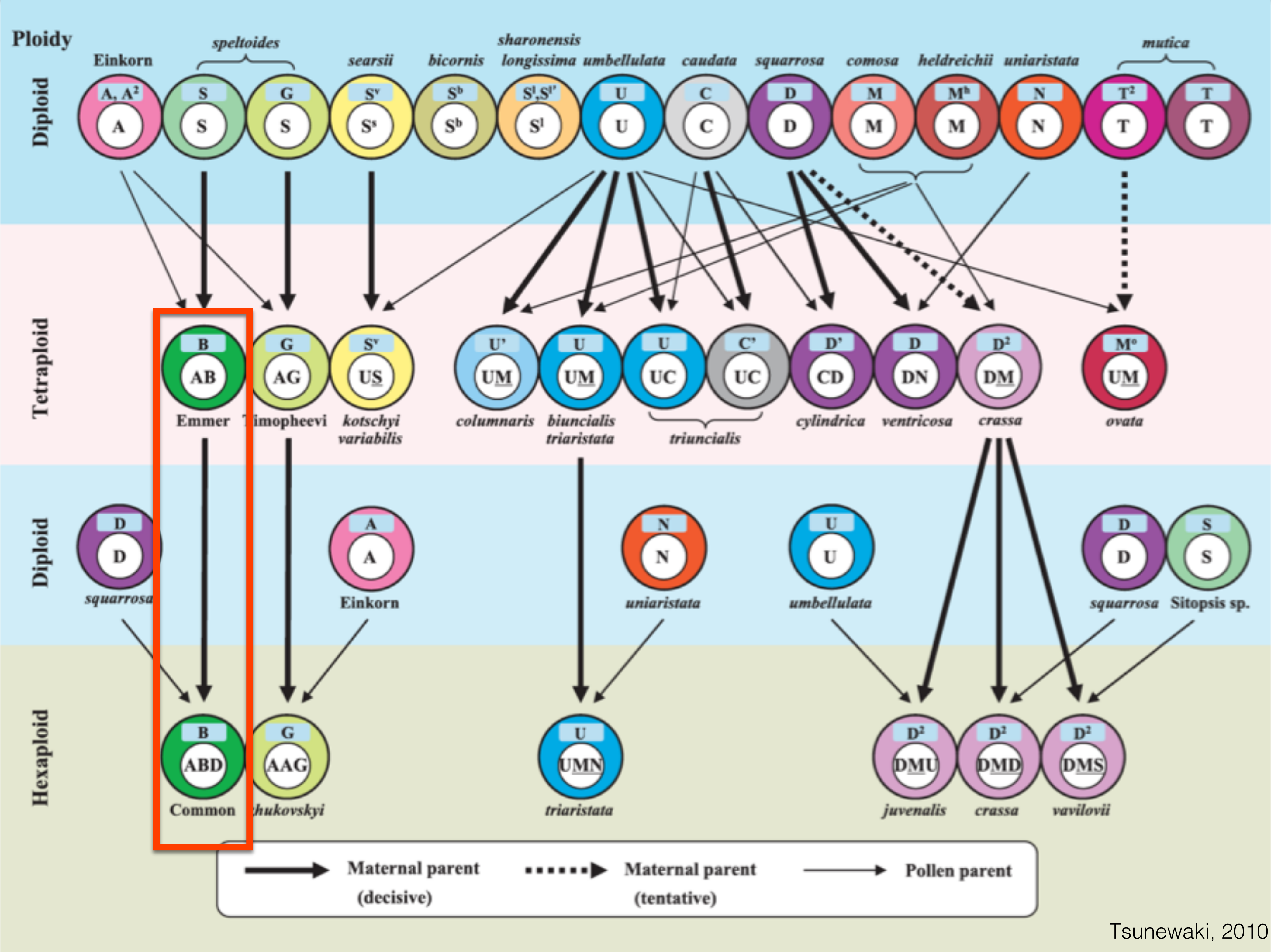
D genome = source of genes for grain quality (glutenins, hardness, etc) environmental adaptability, etc.

Jia et al., (2013) *Nature*: draft of *Ae. tauschii* genome

****Modern breeding utilizes wild species introgression or synthetic hexaploids as sources of genetic diversity to further improve cultivated wheat**

IWGSC (2014) *Science*

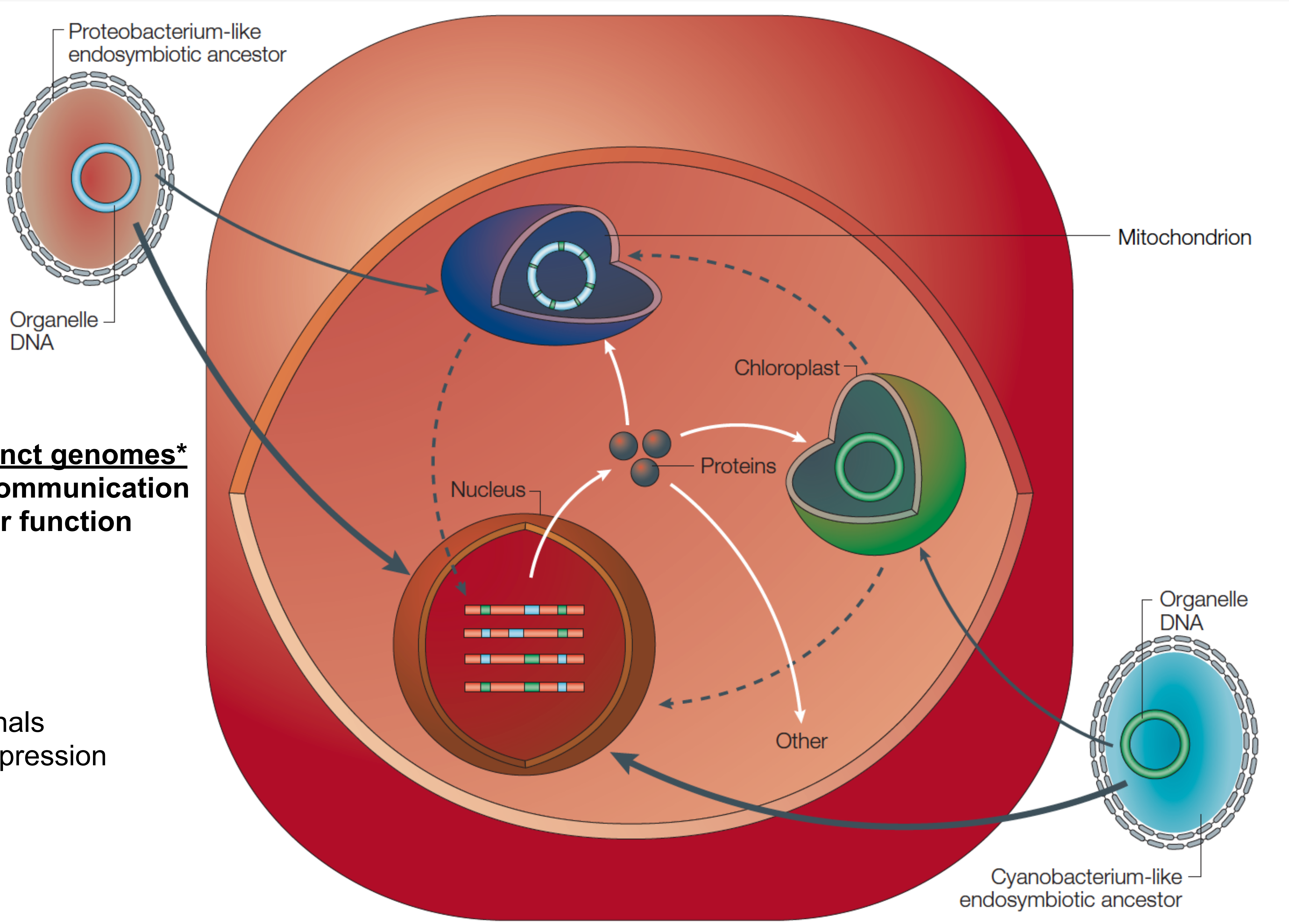
Reduction in cytoplasmic diversity is often overlooked



Tsunewaki, 2010

*Cytoplasmic diversity was reduced to one main type in modern cultivated wheat

Organellar function is critical for development and stress response



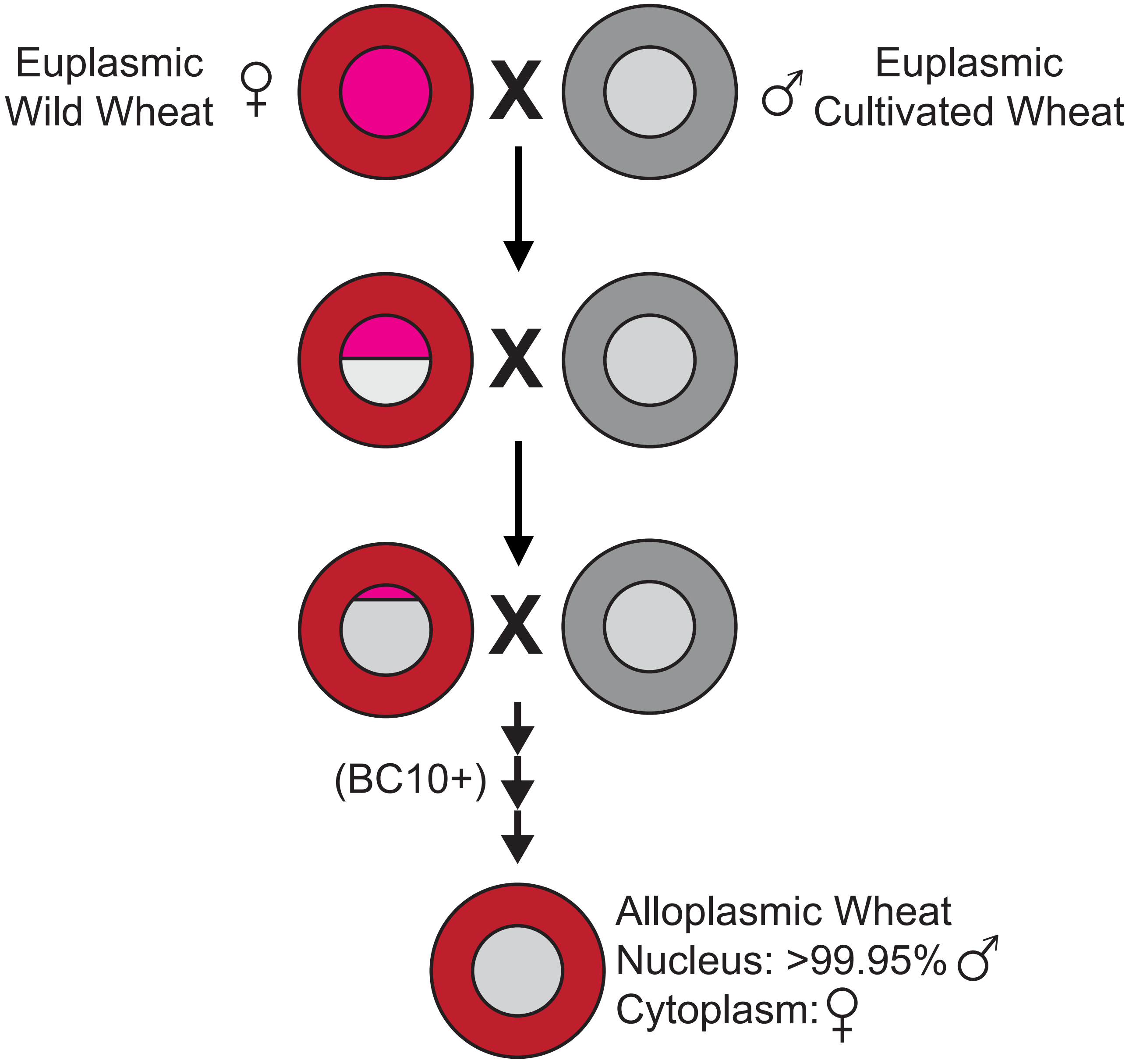
<u>*Co-evolution of distinct genomes*</u>
Requires extensive communication for proper cellular function

- Energy production
- Nutrient sensing
- Photosynthesis
- ROS production
- Integrating stress signals
- Alter nuclear gene expression

Timmis et al. (2004) *Nat Rev Gen*

The wheat alloplasmic collection facilitates the study of organellar diversity and nuclear-cytoplasmic interactions

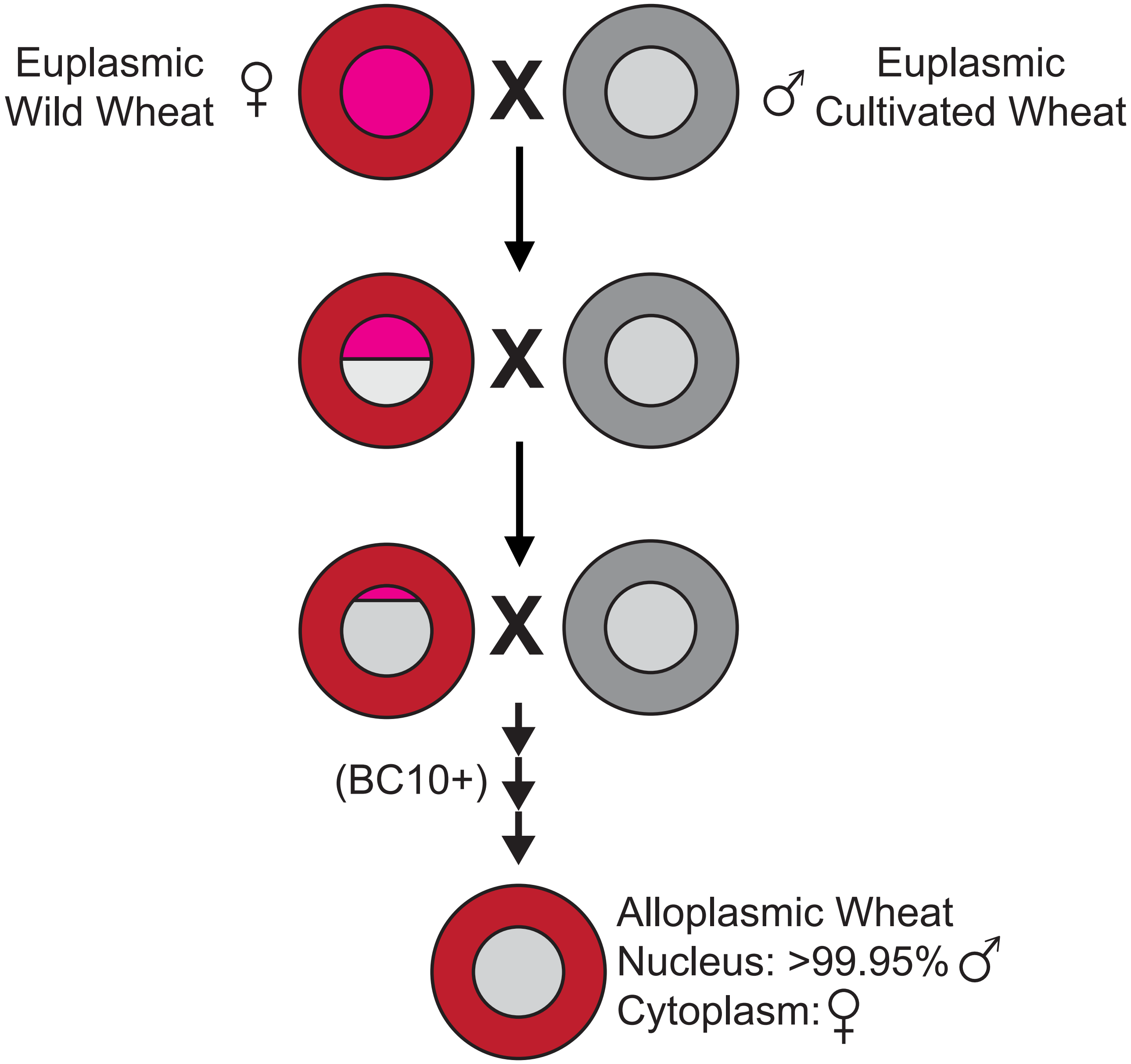
Alloplasmic lines were produced
via recurrent backcrossing schemes



Adapted from: K.L. Liberatore *et al*, (2016) *Free Radical Biol Med*

The wheat alloplasmic collection facilitates the study of organellar diversity and nuclear-cytoplasmic interactions

Alloplasmic lines were produced via recurrent backcrossing schemes



Euplasmic Wild Wheat ♀ X ♂ Euplasmic Cultivated Wheat

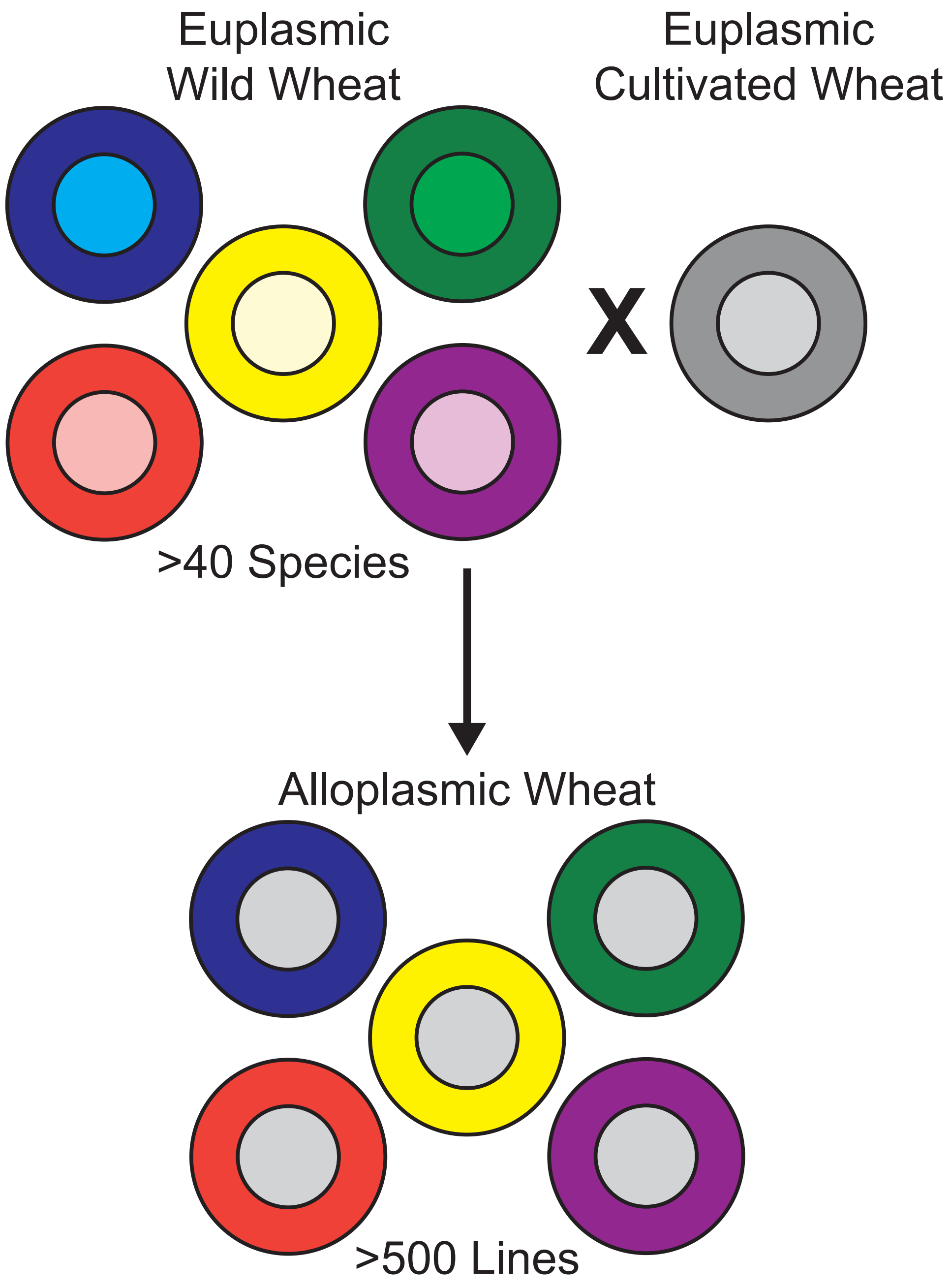
X

X

(BC10+)

Alloplasmic Wheat
Nucleus: >99.95% ♂
Cytoplasm: ♀

Wheat boasts a large alloplasmic collection



Euplasmic Wild Wheat

Euplasmic Cultivated Wheat

X

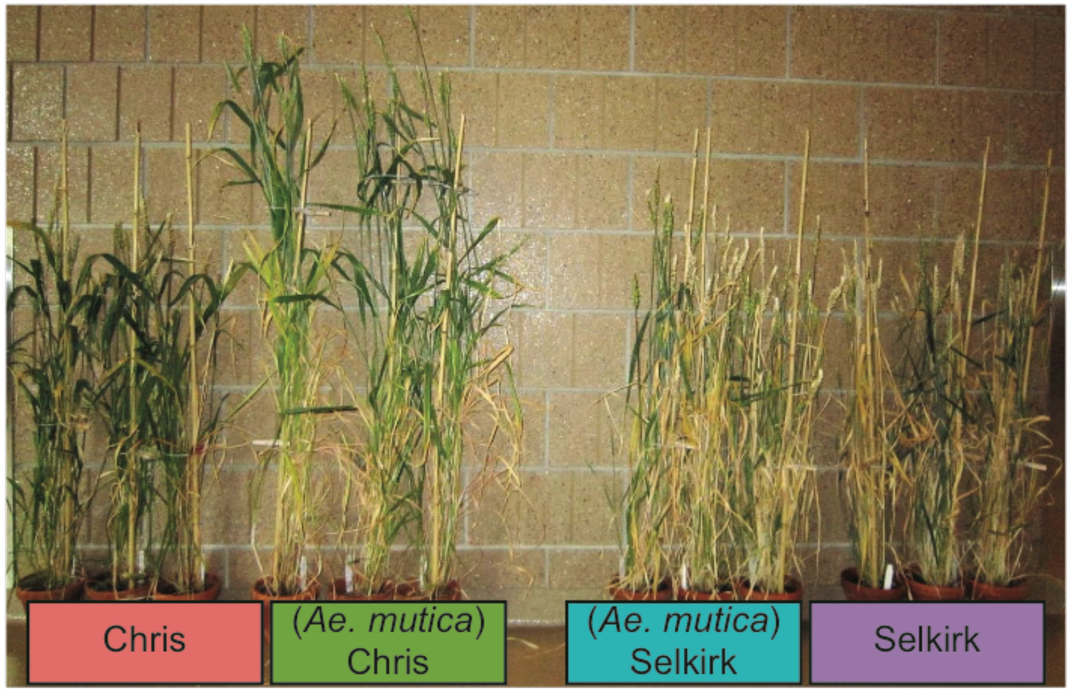
>40 Species

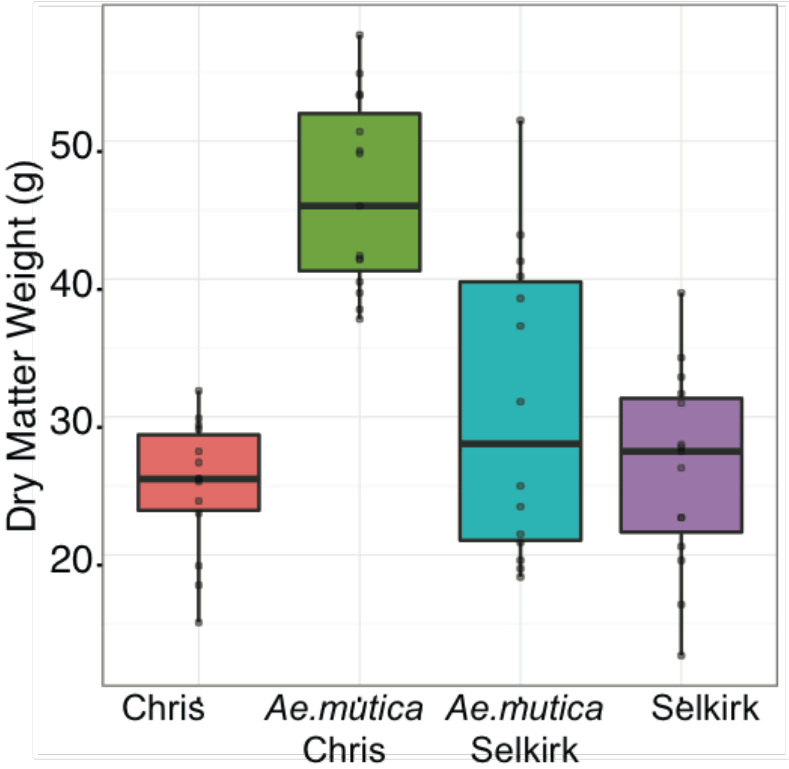
Alloplasmic Wheat

>500 Lines

Adapted from: K.L. Liberatore *et al*, (2016) *Free Radical Biol Med*

Disruption of nuclear-cytoplasmic crosstalk alters wheat developmental processes



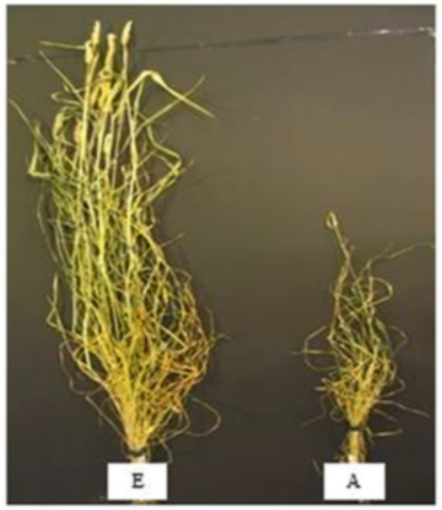


spikelet no.
supernumerary
spikelets

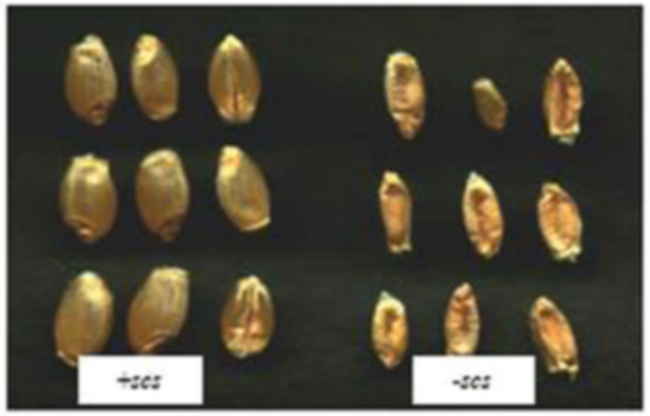
stature

fertility

stature/biomass



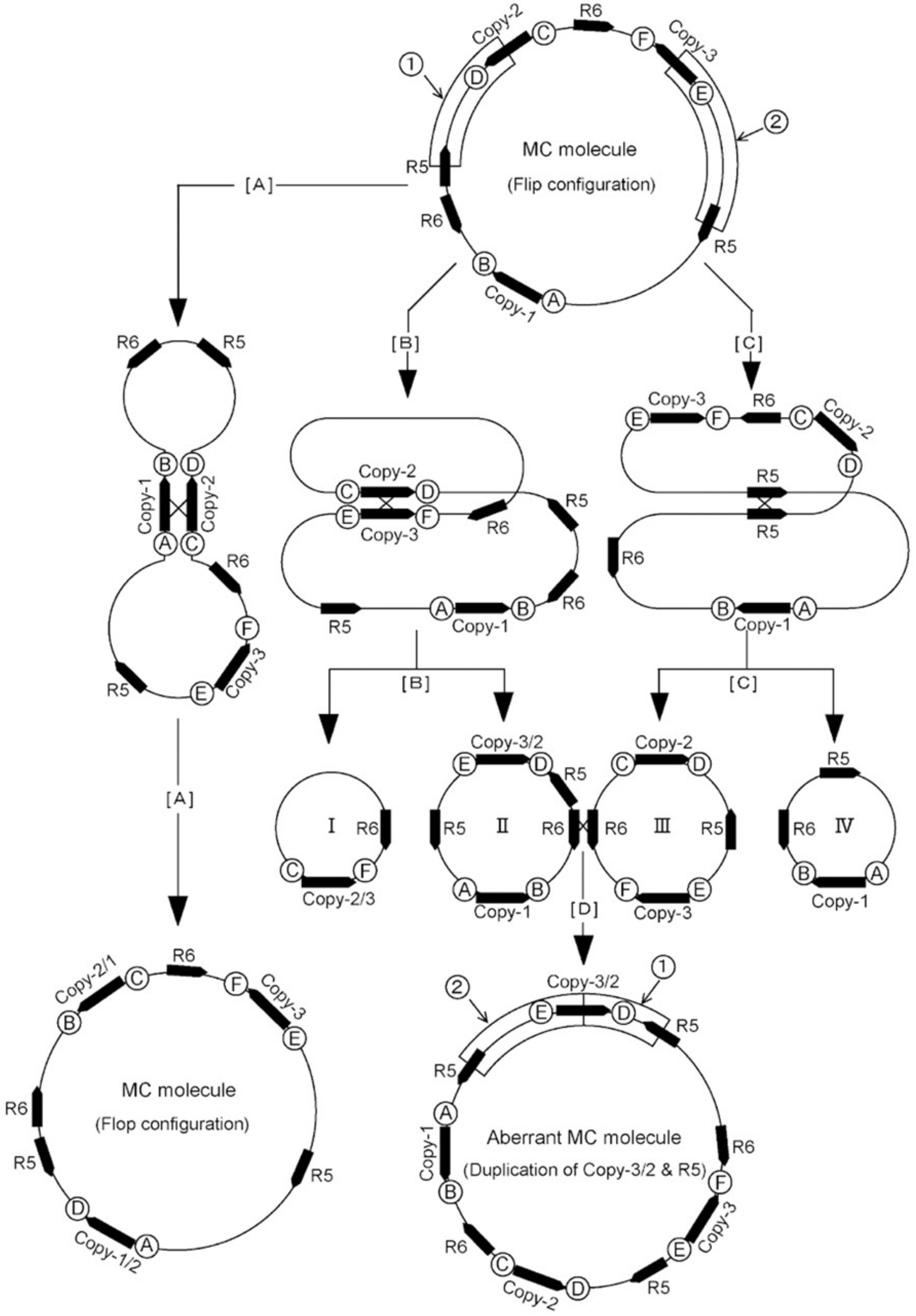
seed size/viability



Adapted from Soltani, A., et al (2016);
Kianian and Kianian (2014);
Liberatore, K.L. et al. (unpublished)

The genetic and structural variation (and dynamics) of the wheat mitochondrial genome are poorly understood

- Published mitochondrial reference genome = ~450kb single “master circle” (by manual stitching)
- Organellar genome rearrangements are facilitated by large repeats
- Cells may contain sub-genomic rings
- Admixtures of genome types may co-exist (heteroplasmy) in a single organelle/cell/tissue type/individual
- Genome types may shift in abundance throughout development or in response to stress (substoichiometric shifting)



Ogihara et al. (2005) *Nuc Acids Res.*

The genetic and structural variation (and dynamics) of the wheat mitochondrial genome are poorly understood

- Published mitochondrial reference genome = ~450kb single "master circle" (by manual stitching)
- Organellar genome rearrangements are facilitated by large repeats
- Cells may contain sub-genomic rings
- Admixtures of genome types may co-exist (heteroplasmy) in a single organelle/cell/tissue type/individual
- Genome types may shift in abundance throughout development or in response to stress (substoichiometric shifting)

Challenges:

Difficult to extract large amounts of pure organellar DNA from limited input tissue

Many wild species and alloplasmic lines are difficult to increase

Ideally would like to sample from individuals

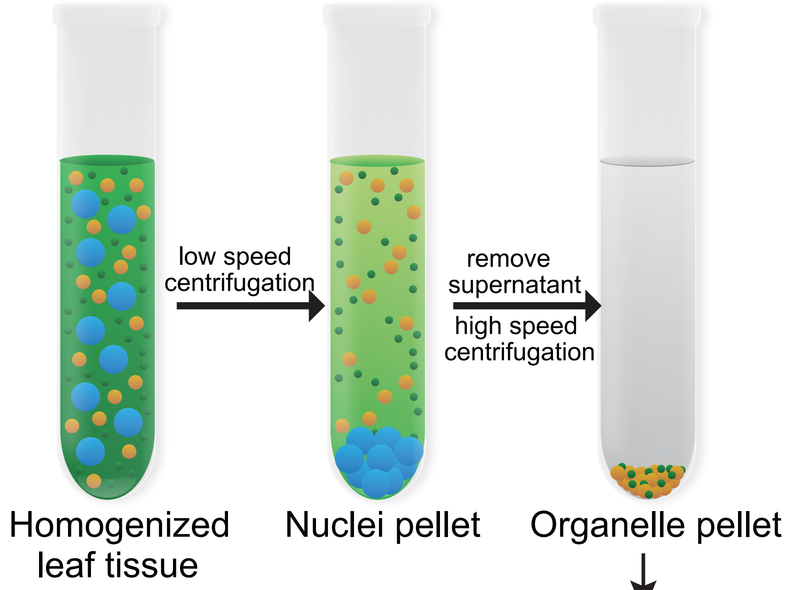
Strategies to enrich for organellar DNA

Differential Centrifugation

Grow seedlings (typically etiolated)

Harvest tissue and process immediately

Differential centrifugation and/or gradients
to isolate intact organelles



DNase treatment
(to eliminate residual nuclear DNA)

Lyse organelles

DNA extraction from
isolated organelles

Purify/concentrate DNA

QC and quantification

Sequencing

Strategies to enrich for organellar DNA

Differential Centrifugation

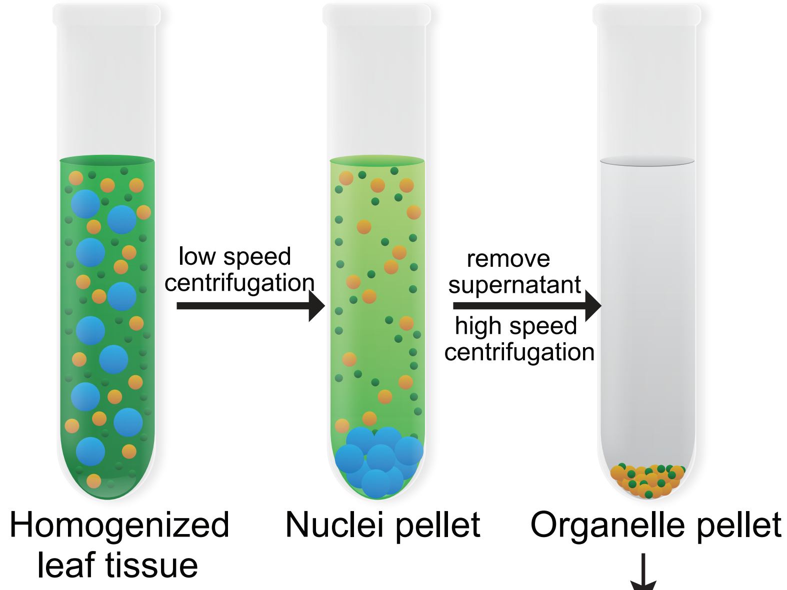
Grow seedlings (typically etiolated)

↓

Harvest tissue and process immediately

↓

Differential centrifugation and/or gradients to isolate intact organelles



↓

DNase treatment
(to eliminate residual nuclear DNA)

↓

Lyse organelles

↓

DNA extraction from isolated organelles

↓

Purify/concentrate DNA

↓

QC and quantification

↓

Sequencing

Increased enrichment for specific organelle

Methyl-Fractionation

Grow seedlings under physiological conditions

↓

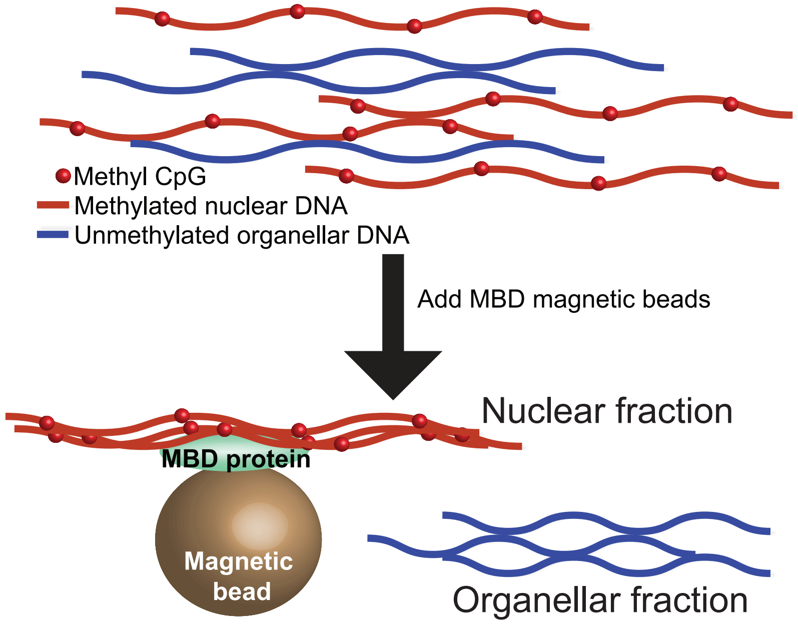
Harvest tissue (Tissue may be stored long-term)

↓

Standard total DNA extraction

↓

Methyl-Binding Domain-mediated fractionation of nuclear and organellar DNA



↓

Purify/concentrate DNA

↓

QC and quantification

↓

Sequencing

1/100 tissue input

tissue does NOT have to be fresh

longer DNA fragments easily maintained

Proportion of each organellar DNA type (plastid vs. mito) maintained closer to native abundances

M.E. Miller and K.L. Liberatore et al (2017);
Adapted from: Yigit et al (2014)

PacBio ultra-low input sequencing results from differential centrifugation (DC) and methyl-fractionation (MF) 20kb library preparations

Triticum aestivum cv. Chinese Spring used for initial *de novo* assemblies to compare to existing Chinese Spring mitochondrial reference genome NC007579.1

Sample	**DC**	**MF-1**	**MF-2**
Total Subreads	59,615	75,966	19,101
Mean subreads length (bp)	6141	5960	5,937
Subread yield (bp)	366,095,715	452,757,360	113,412,212
Mitochondrial %	96.71	8.20	10.4
Chloroplast %	3.21	91.70	89.5
Carrier %	0.08	0.10	0.02

Following protocols of Raley et al (2014 on BioRXiV):

Preparation of next-generation DNA sequencing libraries from ultra-low amounts of input DNA: Application to single-molecule, real-time (SMRT) sequencing on the Pacific Biosciences RS II.

PacBio ultra-low input sequencing results from differential centrifugation (DC) and methyl-fractionation (MF) 20kb library preparations

Triticum aestivum cv. Chinese Spring used for initial *de novo* assemblies to compare to existing Chinese Spring mitochondrial reference genome NC007579.1

Sample	DC	MF-1	MF-2
Total Subreads	59,615	75,966	19,101
Mean subreads length (bp)	6141	5960	5,937
Subread yield (bp)	366,095,715	452,757,360	113,412,212
Mitochondrial %	96.71	8.20	10.4
Chloroplast %	3.21	91.70	89.5
Carrier %	0.08	0.10	0.02

Following protocols of Raley et al (2014 on BioRXiV):
Preparation of next-generation DNA sequencing libraries from ultra-low amounts of input DNA: Application to single-molecule, real-time (SMRT) sequencing on the Pacific Biosciences RS II.

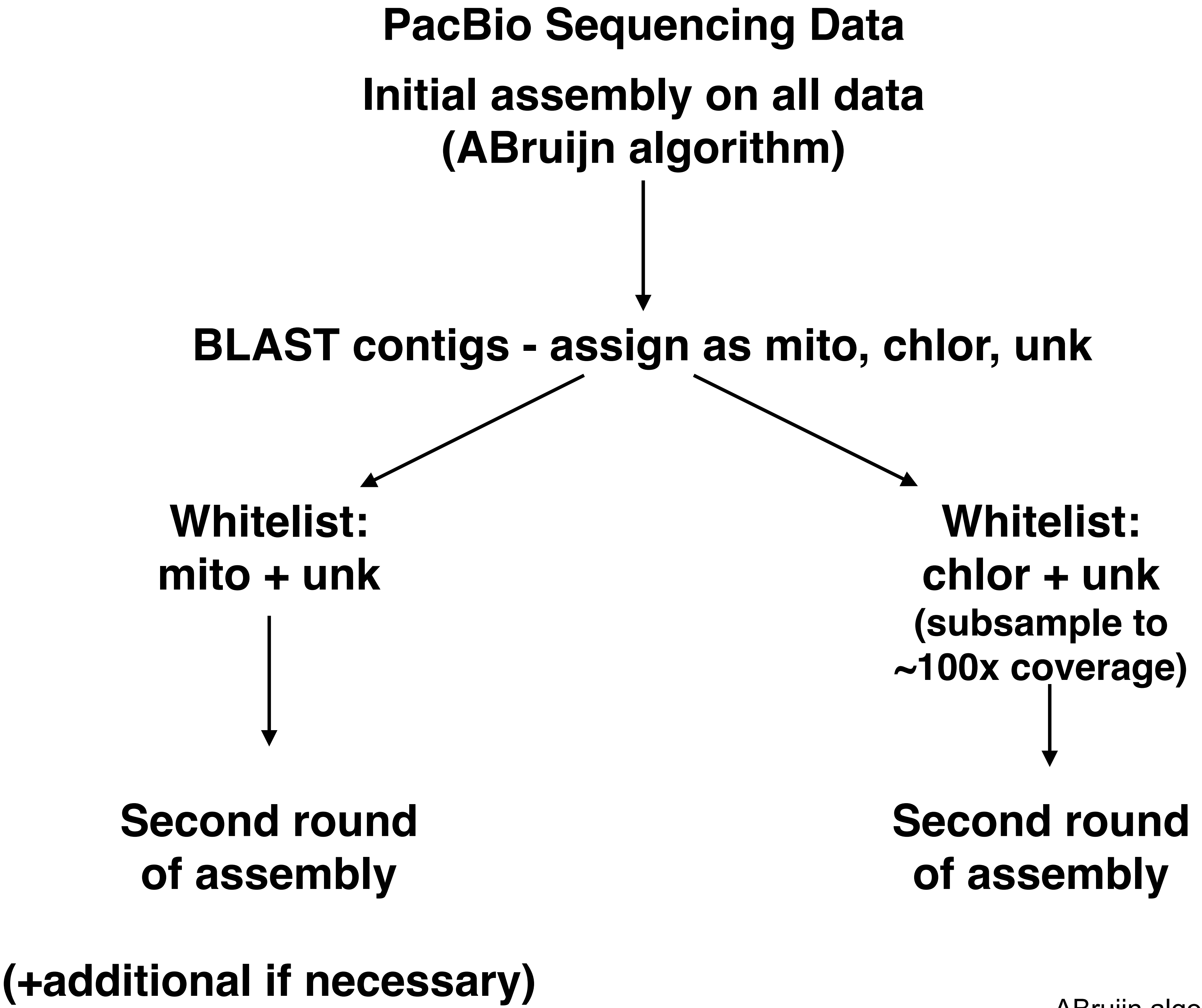
PacBio ultra-low input sequencing results from differential centrifugation (DC) and methyl-fractionation (MF) 20kb library preparations

Triticum aestivum cv. Chinese Spring used for initial *de novo* assemblies to compare to existing Chinese Spring mitochondrial reference genome NC007579.1

Sample	DC	MF-1	MF-2
Total Subreads	59,615	75,966	19,101
Mean subreads length (bp)	6141	5960	5,937
Subread yield (bp)	366,095,715	452,757,360	113,412,212
Mitochondrial %	96.71	8.20	10.4
Chloroplast %	3.21	91.70	89.5
Carrier %	0.08	0.10	0.02

Following protocols of Raley et al (2014 on BioRXiV):
Preparation of next-generation DNA sequencing libraries from ultra-low amounts of input DNA: Application to single-molecule, real-time (SMRT) sequencing on the Pacific Biosciences RS II.

Admixture of mitochondrial and plastid sequences are separated computationally during *de novo* assembly



PacBio Sequencing Data

Initial assembly on all data
(ABruijn algorithm)

BLAST contigs - assign as mito, chlor, unk

Whitelist:
mito + unk

Whitelist:
chlor + unk
(subsample to
~100x coverage)

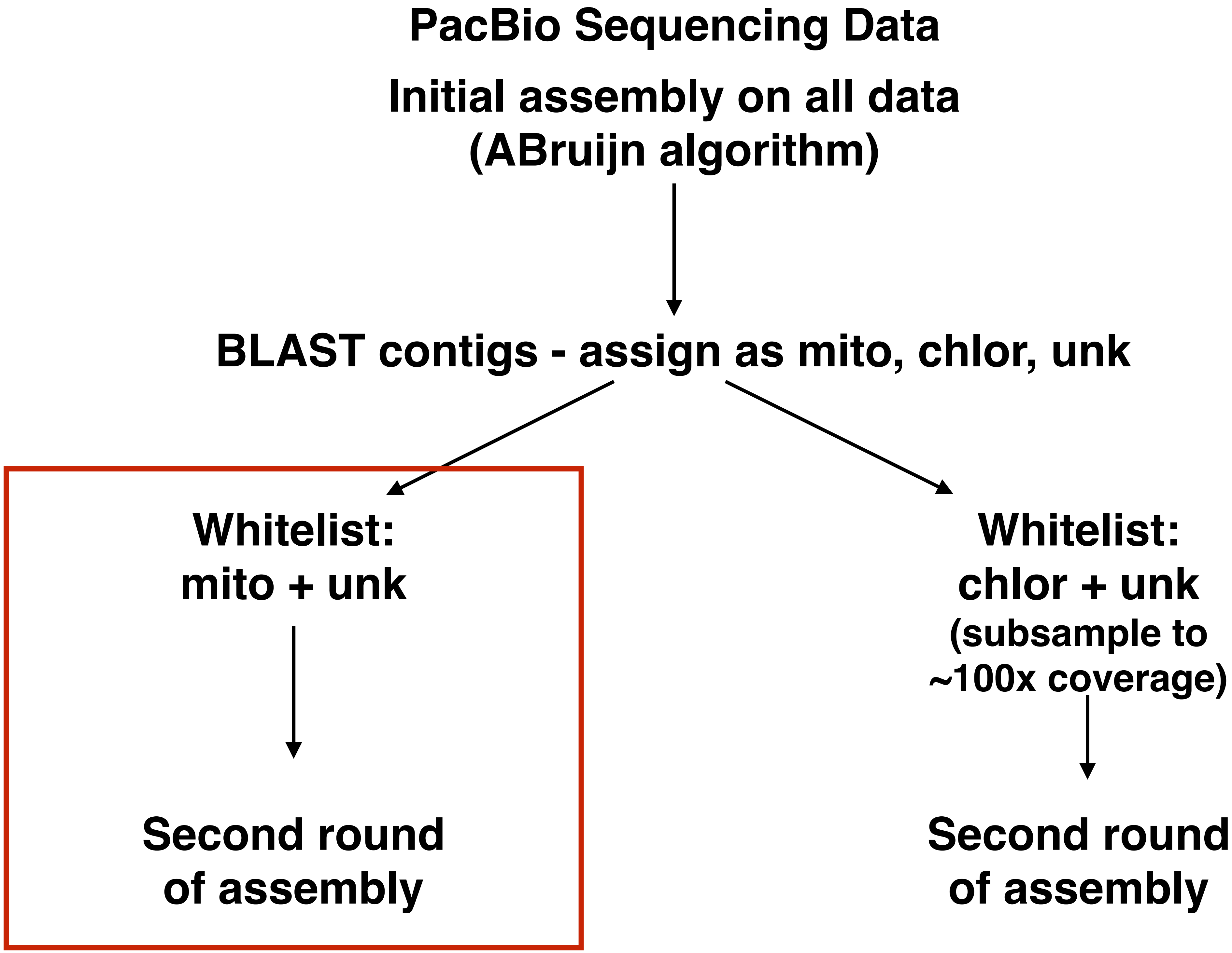
Second round
of assembly

(+additional if necessary)

Second round
of assembly

ABruijn algorithm: Lin et al. (2016) *PNAS*

Admixture of mitochondrial and plastid sequences are separated computationally during *de novo* assembly



PacBio Sequencing Data

Initial assembly on all data
(ABruijn algorithm)

BLAST contigs - assign as mito, chlor, unk

Whitelist:
mito + unk

Second round
of assembly

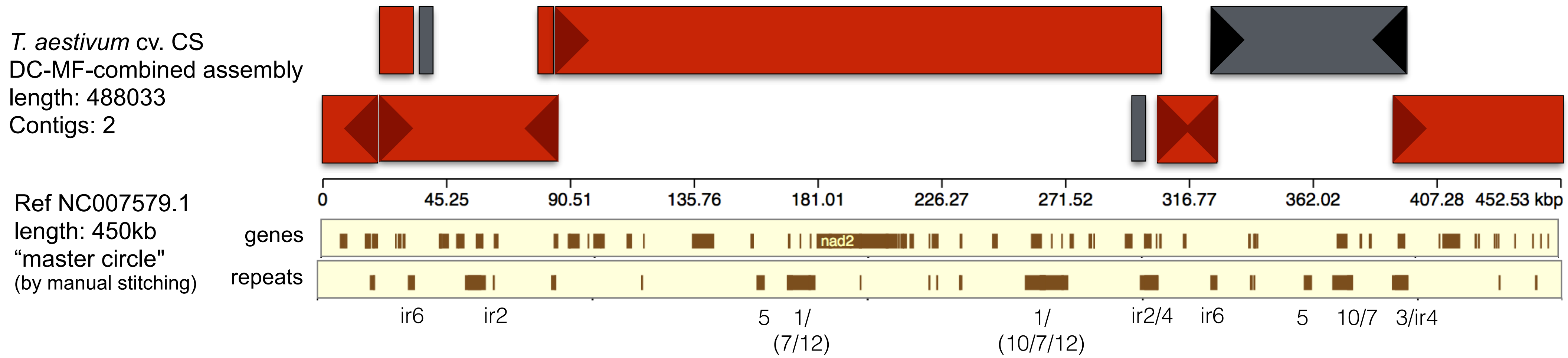
Whitelist:
chlor + unk
(subsample to
~100x coverage)

Second round
of assembly

(+additional if necessary)

ABruijn algorithm: Lin et al. (2016) *PNAS*

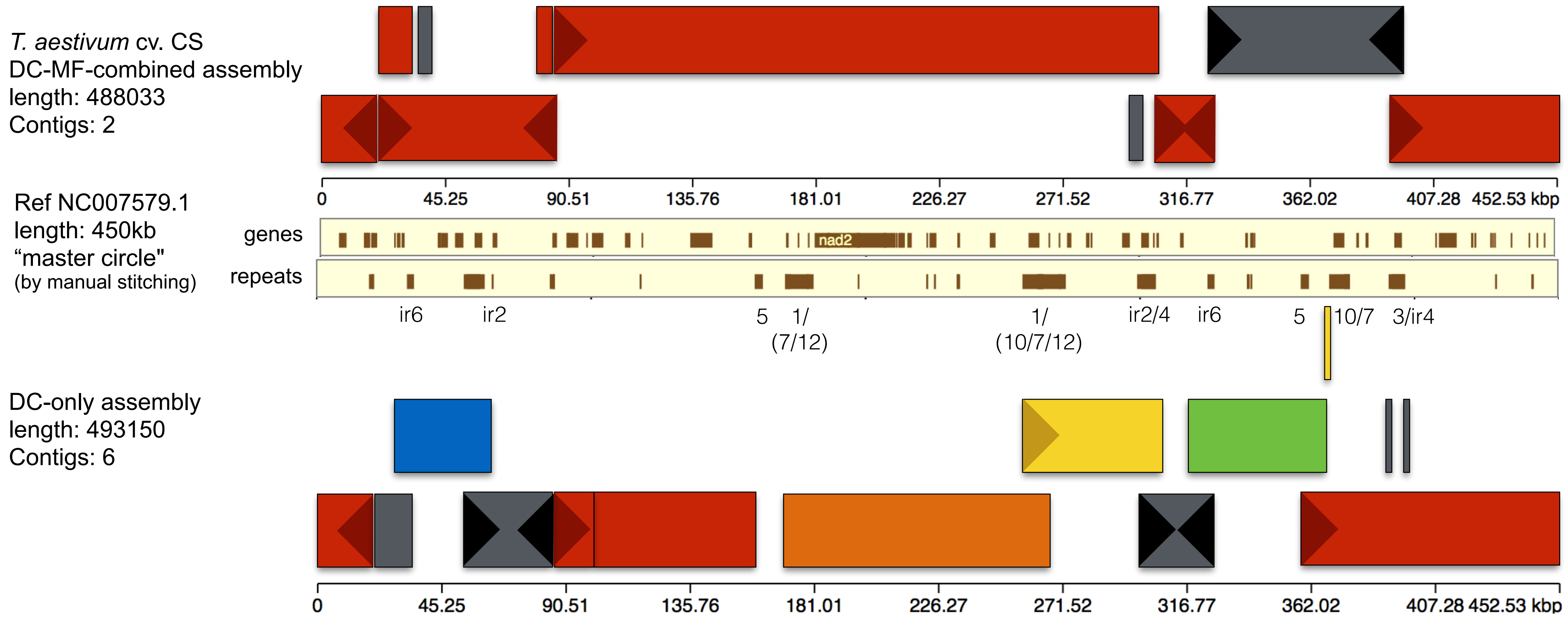
The DC and MF combined data set assembly covers the full published CS mitochondria ref genome NC007579.1 with an average of 500x coverage



Red bars =
longest contig
in each assembly

Assembly	average coverage	contig no.	genome size	largest contig	longest alignment	% ref. genome	total gene no.	CDS no.	repeat no.
Reference NC007579.1	—	master circle (manual stitching)	452528	—	—	—	81	47	35
combined data	>500x	2	488090	408831	221719	100	81	47	35

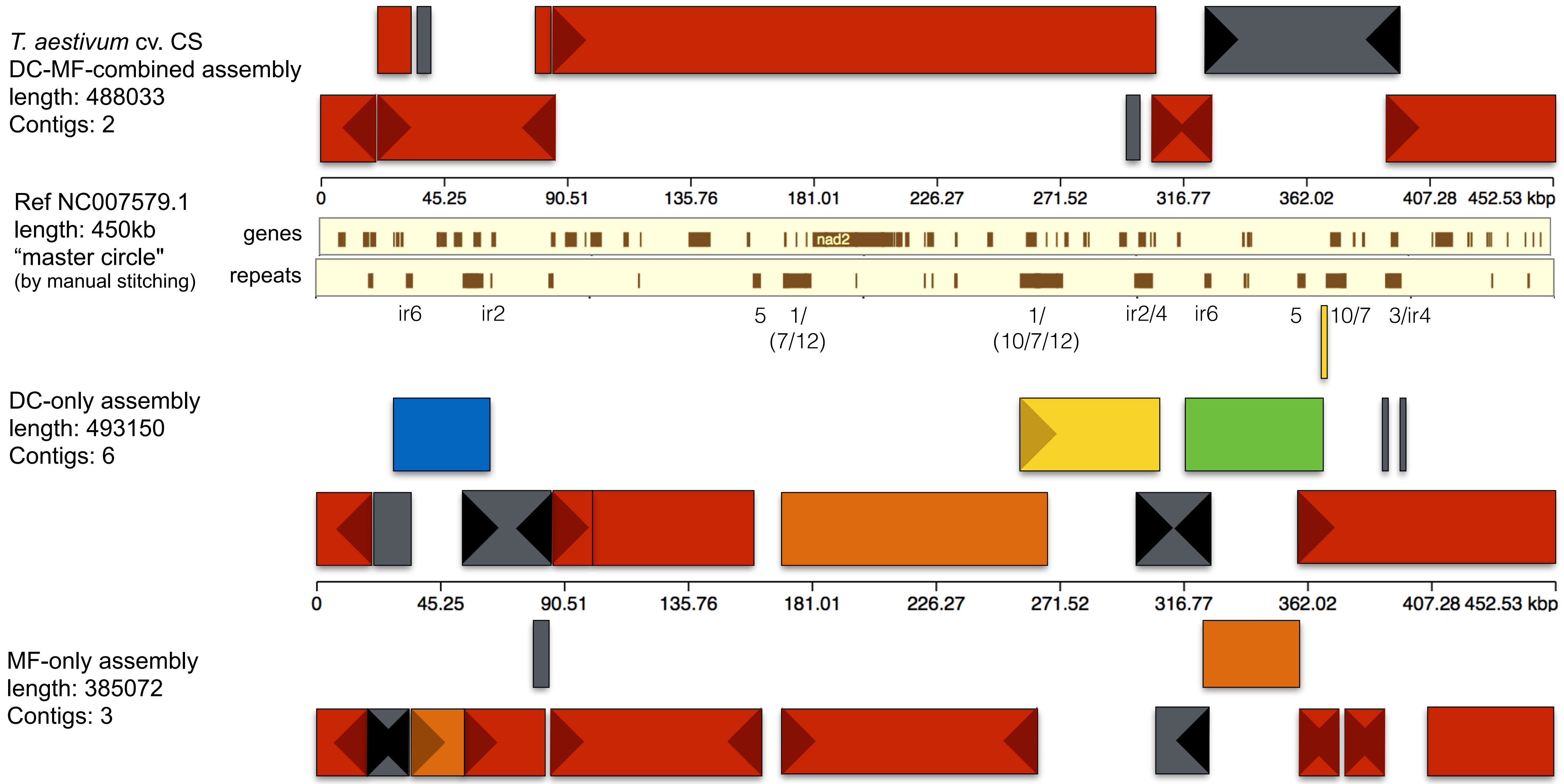
The DC dataset alone nearly covers the full mitochondria reference genome but remains fragmented at repeats with an average of >400x coverage



Red bars = longest contig in each assembly

Assembly	average coverage	contig no.	genome size	largest contig	longest alignment	% ref. genome	total gene no.	CDS no.	repeat no.
Reference NC007579.1	—	master circle (manual stitching)	452528	—	—	—	81	47	35
combined data	>500x	2	488090	408831	221719	100	81	47	35
DC only	~450x	6	493150	187711	114399	97.57	80 + 1 partial	46 + 1 partial	34 + 0 partial

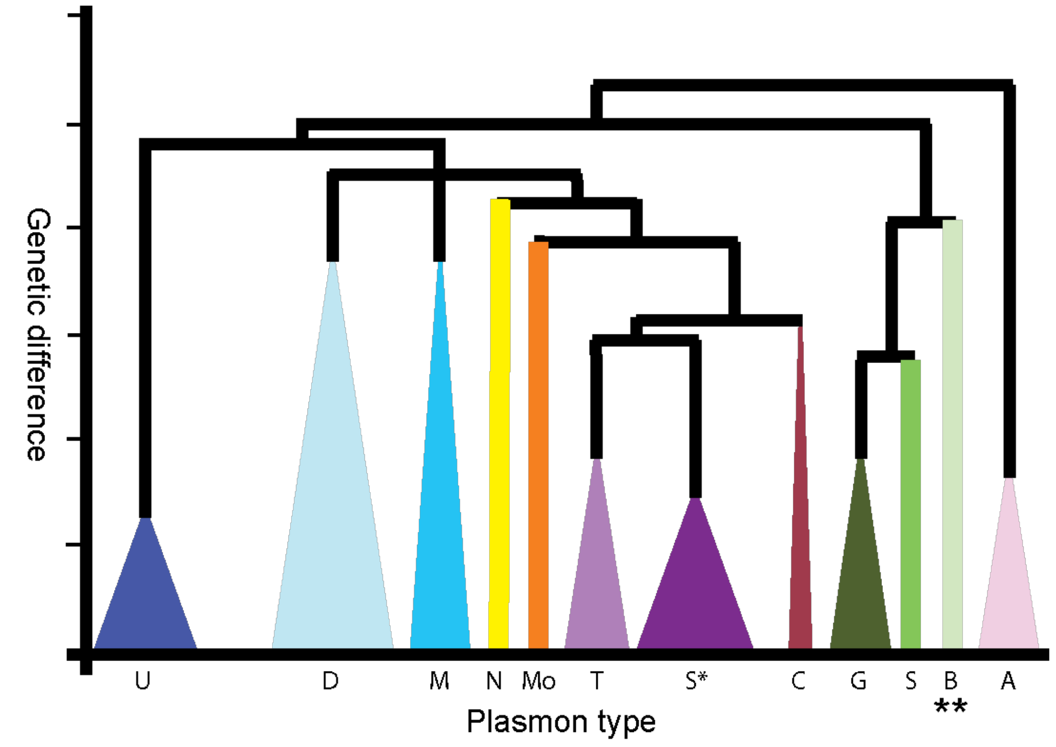
The MF data alone covers a majority of the mitochondrial genome with an average of 60x coverage



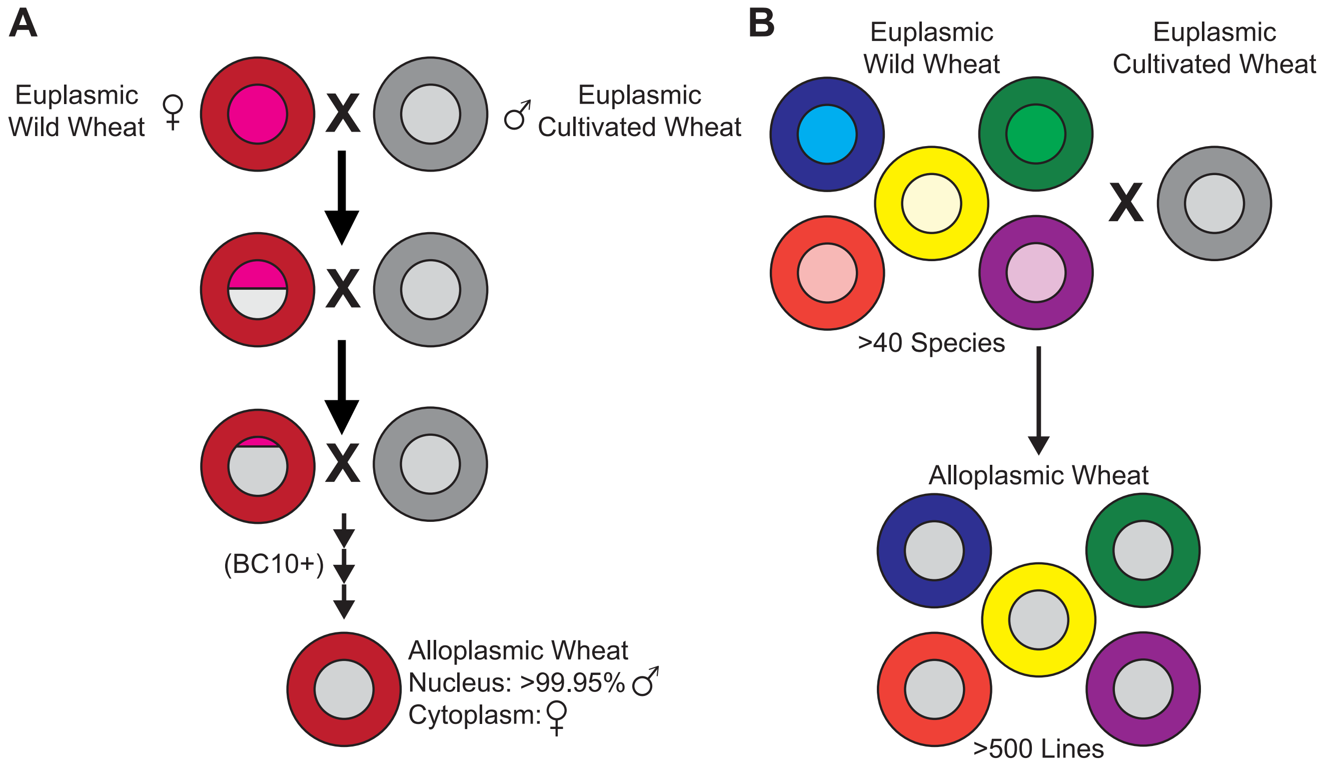
Red bars = longest contig in each assembly

Assembly	average coverage	contig no.	genome size	largest contig	longest alignment	% ref. genome	total gene no.	CDS no.	repeat no.
Reference NC007579.1	—	master circle (manual stitching)	452528	—	—	—	81	47	35
combined data	>500x	2	488090	408831	221719	100	81	47	35
DC only	~450x	6	493150	187711	114399	97.57	80 + 1 partial	46 + 1 partial	34 + 0 partial
MF only	~60x	3	385072	292460	93546	83.66	65 + 1 partial	41 + 0 partial	24 + 3 partial

Expanding wheat organellar genomic resources



Adapted from: K. Tsunewaki et al (2009)



Adapted from: K.L. Liberatore *et al*, (2016) *Free Radical Biol Med*

Species	Cytoplasm "type"
Ae. squarrosa	D
Ae. columnaris	U'
Ae. ventricosa	D
T. durum	B
T. dicoccoides	B
Ae. caudata	C
Ae. crassa	D2
Ae comosa	M
Ae. ovata	Mo
Ae uniaristata	N
Ae. speltoides	S/G
Ae. sharonensis	S1/S1'
Ae. mutica	T2/T
Ae. biuncialis	U

Alloplasmics	Cytoplasm "type"
(*Ae. squarrosa*) Chinese Spring	D
(*Ae. columnaris*) Chinese Spring	U'
(*Ae. ventricosa*) Chinese Spring	D

*All of these samples, plus an additional 75 selected wild species and alloplasmic lines were sequenced by Illumina

PacBio ultra-low input sequencing results from MF 20kb library preparations of wild relatives and select alloplasmic lines

Genotype	Total Subreads	Avg. ROI (bp)	Total Bases (Mbp)	Mito Yield (Mbp)	Mito %	Mito Avg Coverage
Ae. squarrosa	106,566	6,798	642.4	36.3	7.71%	80.3
(Ae. squarrosa) Chinese Spring Allo	64,486	6,055	320.0	16.6	7.13%	36.8
Ae. columnaris	15,550	5,944	86	4.9	7.93%	10.9
Ae. columnaris	30,496	6,081	185.5	13.9	10.7%	30.9
(Ae. columnaris) Chinese Spring Allo	180,988	5,663	861	34.9	6.75%	77.3
Ae. ventricosa	225,574	6,104	1,181	55.6	7.05%	122.9
(Ae. ventricosa) Chinese Spring Allo	155,453	6,135	832.8	30.4	5.32%	67.2
T. dicoccoides	125,214	6,365	797.0	102.7	19.1%	227.0
Ae. caudata	271,167	6,522	1,511	112.4	12.33%	248.4
Ae. crassa	166,795	6,843	921.3	57.0	9.15%	126.0
Ae. comosa	356,335	7,302	2,052	127.7	11.11%	282.3
Ae. ovata	133,211	5,948	708.7	45.5	9.54%	100.6
Ae. uniaristata	172,385	6,233	1,074	119.1	15.8%	263.3
Ae. speltoides	231,941	7,063	1,485	78.4	8.42%	173.3
Ae. sharonensis	94,611	5,443	477.7	35.8	11.04%	79.3
Ae. mutica	168,834	6,741	1,011	27.8	3.85%	61.6
Ae. biuncialis	55,510	5,709	280.3	24.6	11.88%	54.4
T. durum	58,470	5,827	340.7	27.1	10.5%	60.0

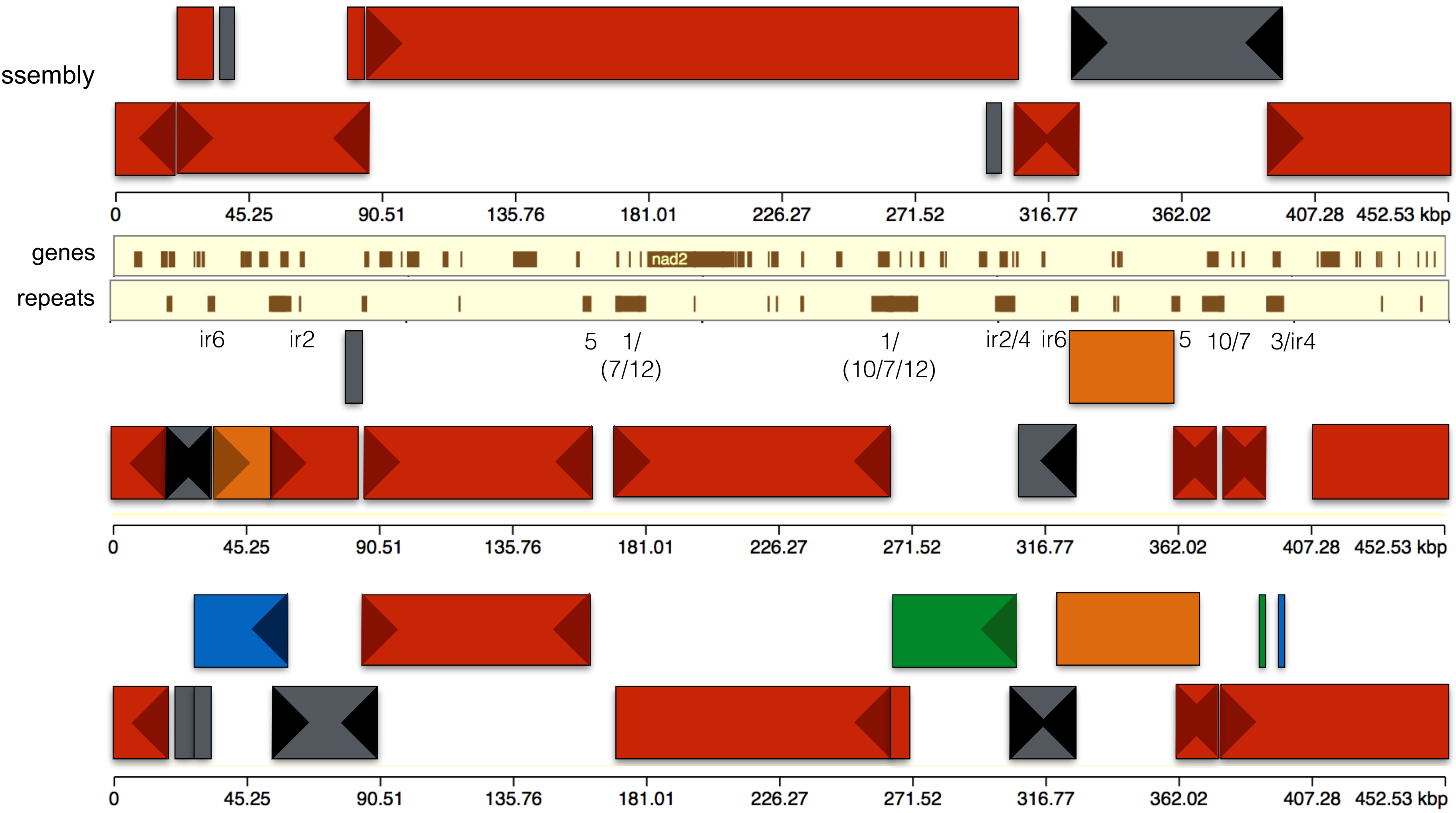
T. dicoccoides mitochondria draft assembly produced with the MF procedure (Single PacBio RSII SMRT Cell yielded ~150x average coverage)

T. aestivum cv. CS
DC-MF-combined assembly
length: 488033
Contigs: 2

Ref NC007579.1
length: 450kb
“master circle"
(by manual stitching)

T. aestivum cv. CS
MF-only assembly
length: 385072
Contigs: 3

T. dicoccoides
length: 473471
Contigs: 5



Red bars = longest contig in each assembly

Assembly	average coverage	contig no.	genome size	largest contig	longest alignment	% ref. genome	total gene no.	CDS no.	repeat no.
Reference NC007579.1	—	master circle (manual stitching)	452528	—	—	—	81	47	35
combined data	>500x	2	488090	408831	221719	100	81	47	35
MF only	~60x	3	385072	292460	93546	83.66	65 + 1 partial	41 + 0 partial	24 + 3 partial
T. dicoccoides	~150x	5	473471	286927	95674	97.7	79 + 0 partial	46 + 0 partial	30 + 3 partial

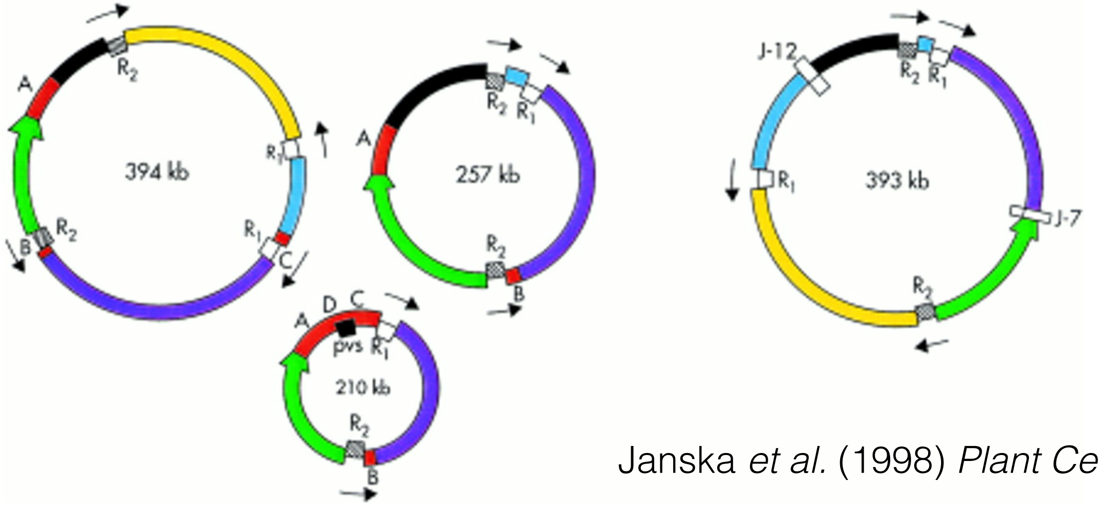
A few direct links between organellar genome variants and altered plant development

Cytoplasmic Male Sterility (CMS)

Common Bean CMS

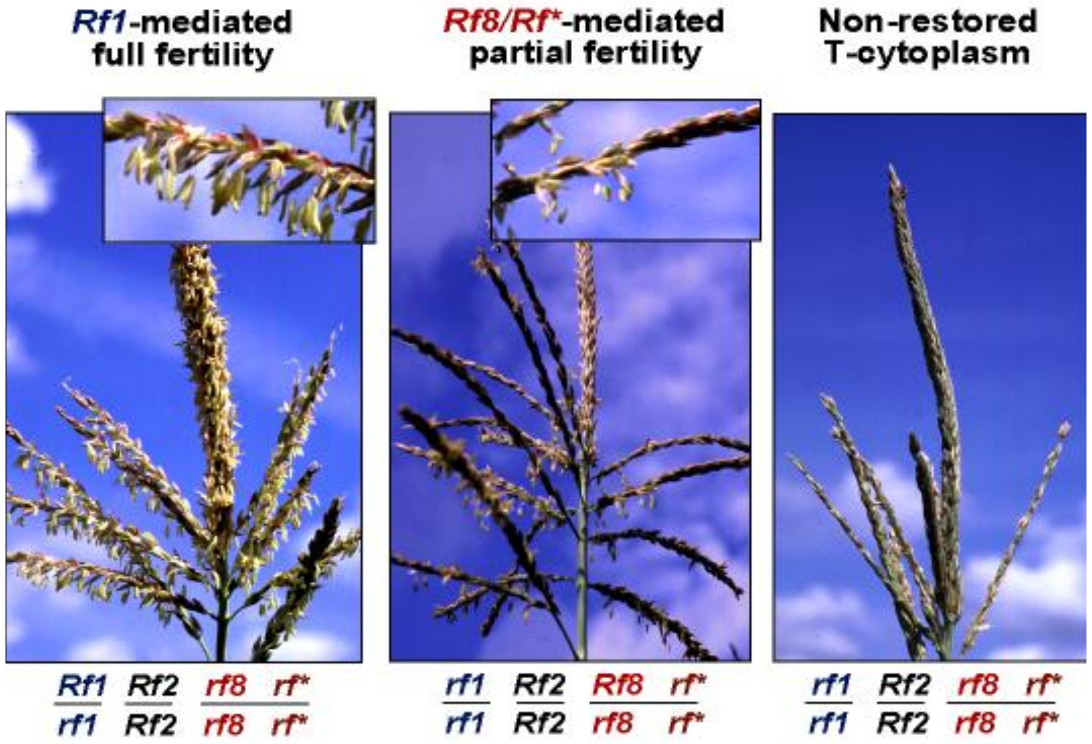
CMS (sterile) line

Fertile line



Janska *et al.* (1998) *Plant Cell*

Maize CMS


***Rf1*-mediated full fertility**

***Rf8/Rf**-mediated partial fertility**

Non-restored T-cytoplasm

Dill *et al.* (1997) *Genetics*

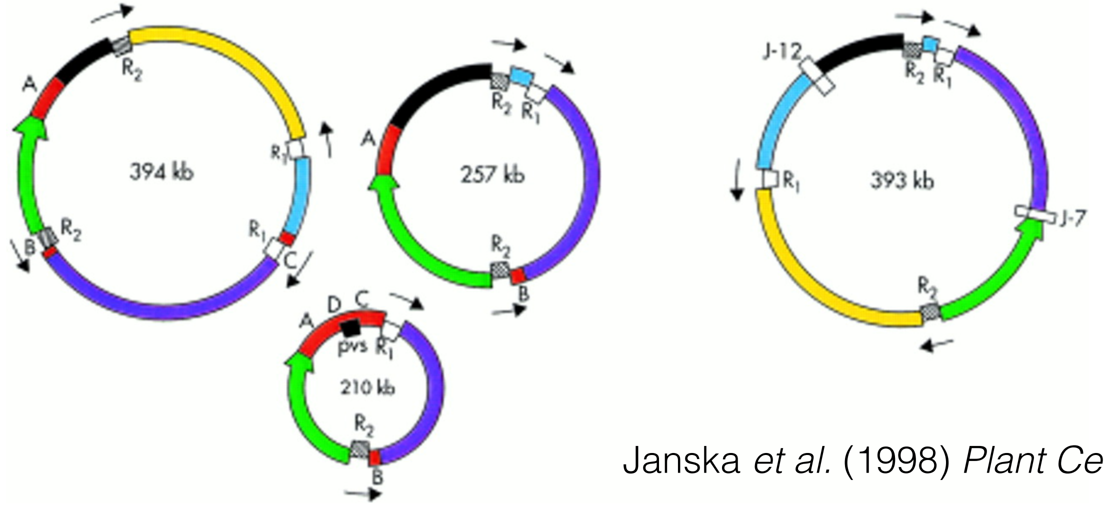
Functional implications of specific wheat organellar variants?

Cytoplasmic Male Sterility (CMS)

Common Bean CMS

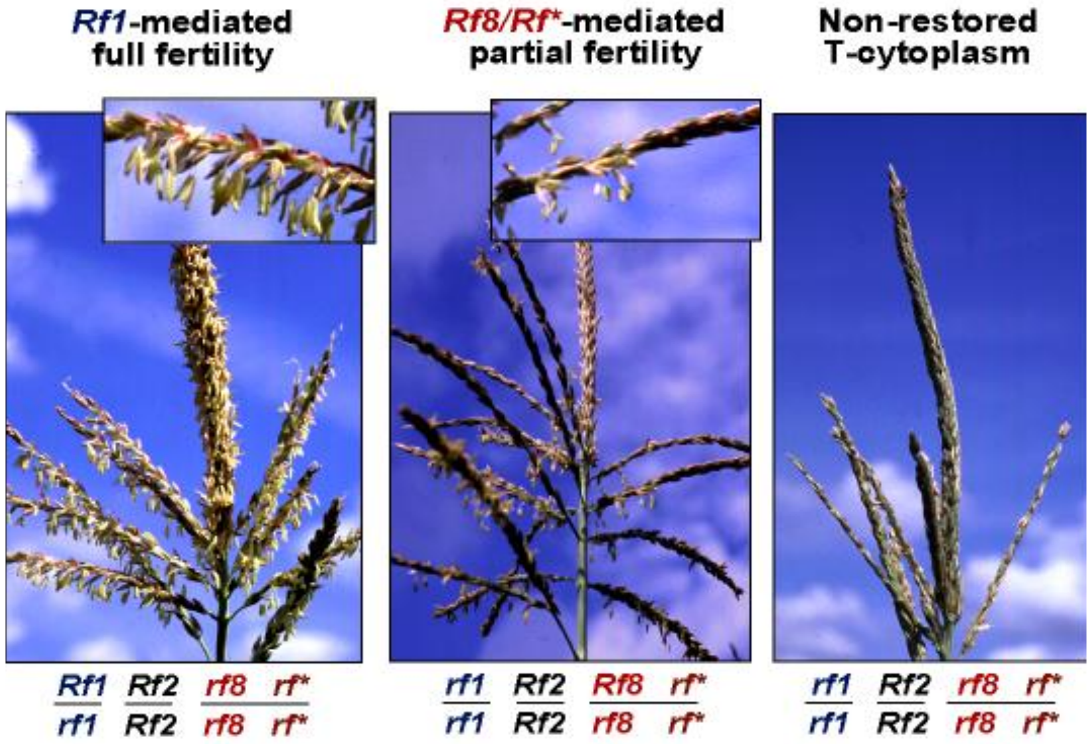
CMS (sterile) line

Fertile line



Janska *et al.* (1998) *Plant Cell*

Maize CMS

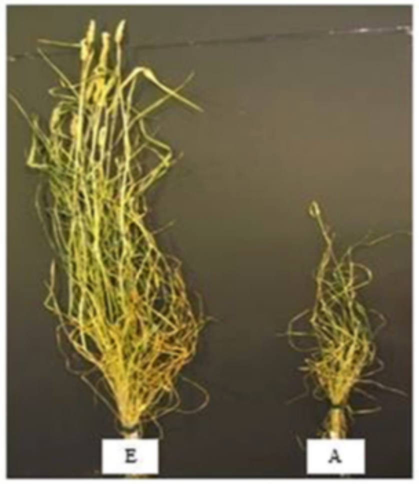


Dill *et al.* (1997) *Genetics*

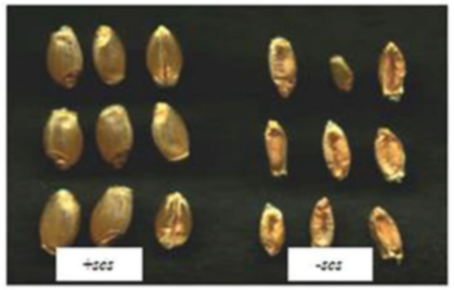
Wheat mitochondrial characterization in progress:

- Molecular validation of contigs and putative sub-genomic rings
- Cross-species comparative analyses
- Investigation of heteroplasmy

spikelet no. stature/biomass



seed size/viability



Kianian & Kianian (2014)

Thank you!

USDA

UNIVERSITY OF MINNESOTA
Driven to Discover®

Shahryar F. Kianian
Marisa E. Miller
Roger Caspers

UMN Genomics Center staff

Minnesota Supercomputing Institute

Kevin A.T. Silverstein

leidos
Leidos Biomedical Research, Inc.

Castle Raley
Bao Tran
Jack Chen

PACBIO®

Laura Nolden
Jenny Alfrey

NEW ENGLAND BioLabs®

Erbay Yigit
Isabel Gautreau
Rick Morgan

Funding:

NSF

PGRP: EAGER
IOS #1361554

USDA