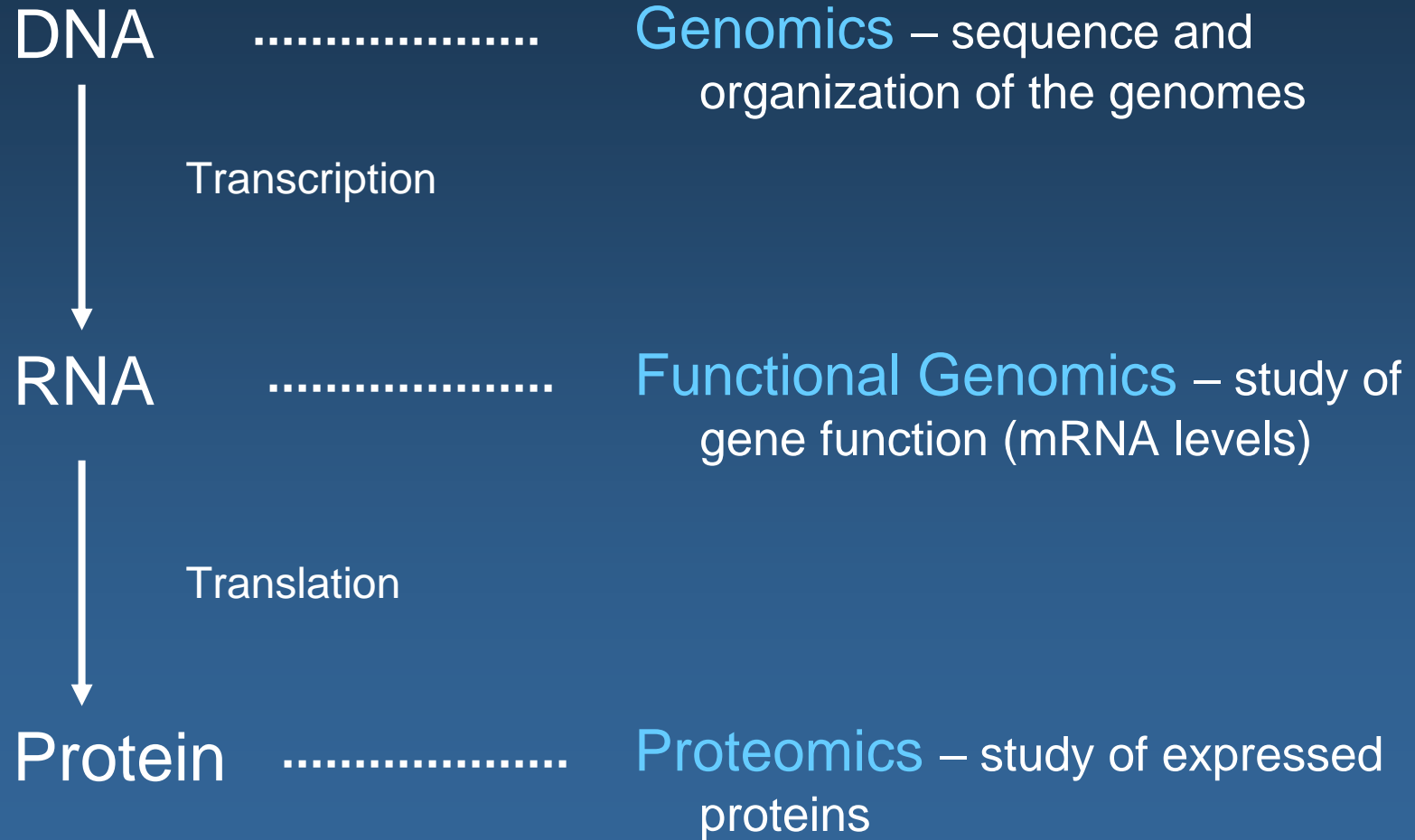


A wide-angle photograph of a field of yellow Brassica flowers, likely rapeseed, stretching to the horizon. The sky is a deep blue with scattered white and grey clouds. The text is overlaid on the upper half of the image.

Developing a *Brassica* microarray resource

Andrew Sharpe
AAFC Saskatoon Research Centre

Levels of genome analysis



Functional genomics

- Gene expression analysis
 - measures the levels of mRNA in cells
 - study of genes expressed by a specific cell or group of cells and the changes in their expression patterns during development, disease, or environmental exposure



Gene expression analysis

- Northern blotting
- dot blots
- quantitative RT-PCR

Measurement of levels of mRNA, one gene at a time



Gene expression analysis

- EST - Expressed Sequence Tags
- SAGE - Serial Analysis of Gene Expression
- MPSS - Massively Parallel Signature Sequencing
- Hybridization to microarrays

Measurement of levels of mRNA, 10,000s genes at a time



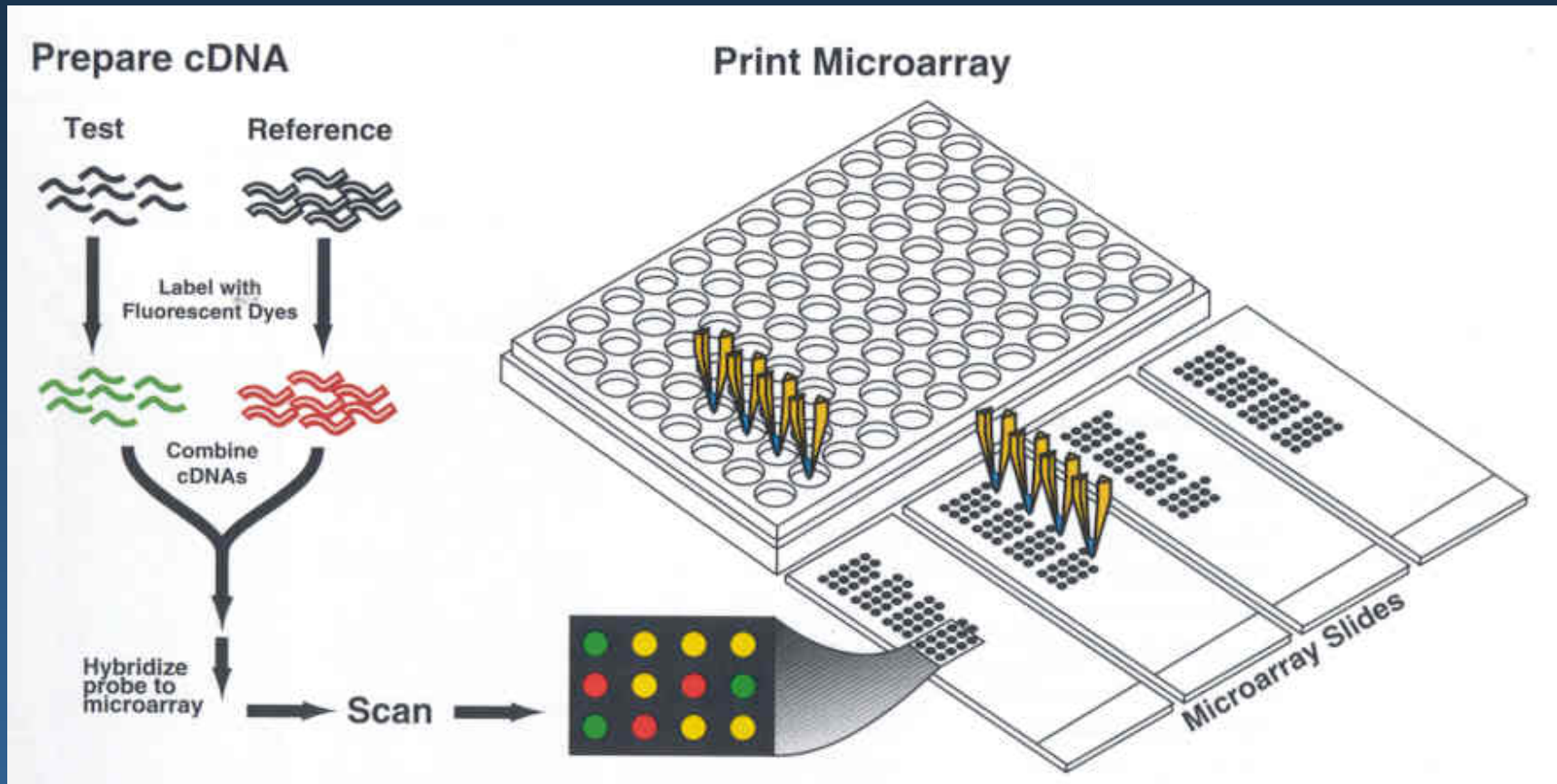
Microarrays

Spotted DNA microarrays:

- DNA - Portions of genes or short oligonucleotides
- DNAs are spotted by a robot on glass slides
- 10,000s spots can be arrayed



Microarrays



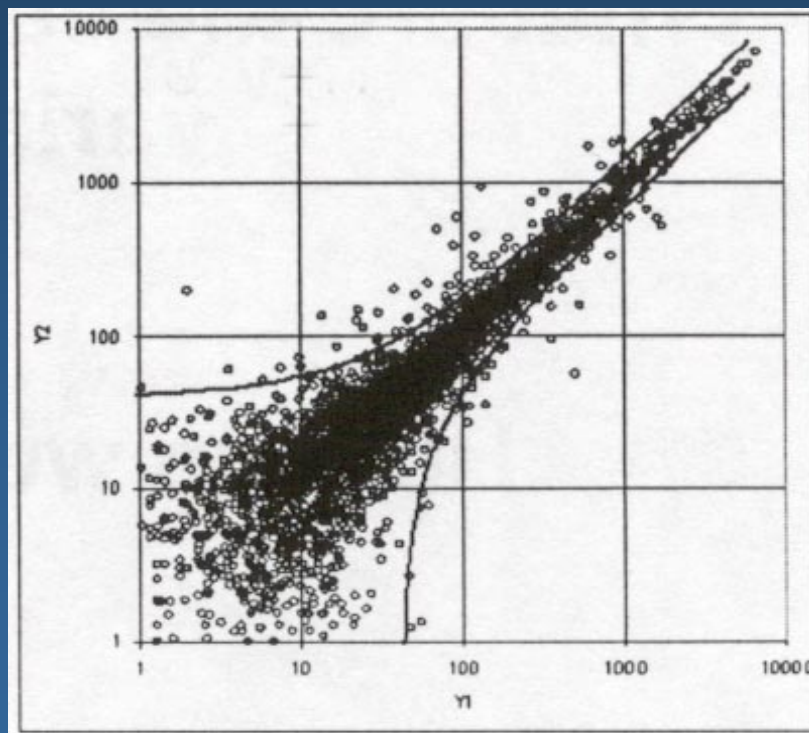
- RNA extracted from test and reference samples
- RNAs labelled with different dyes – cDNA target



Microarrays

Spotted DNA microarrays

- Difference in the ratio of hybridization between samples is measured for all DNA probes
- Over and under expressing genes are identified in test sample



Microarrays

Key issues:

- Measures only relative and not absolute levels of gene transcription
- Replicate experiments and good experimental design
- Analysis of complex data sets to extract biological significance (understanding the results)
- Information technology issues (managing large data sets and tracking samples)



Microarrays at AAFC

Global *Arabidopsis* / *Brassica* transcription profiling

- Range of different applications:
 - abiotic and biotic stresses
 - specific expression of gene family members
 - effects of ploidy on gene expression
 - interesting lines – mutants, NILs
- Current expression experiments performed using an *Arabidopsis* 26,000 70mer oligonucleotide array
 - Considerations for *Brassica* work:
 - Relatedness of *Brassica* and *Arabidopsis*
 - *Brassica* specific genes



Amphidiploid *B. napus* ($2n=38$)
Important crop species

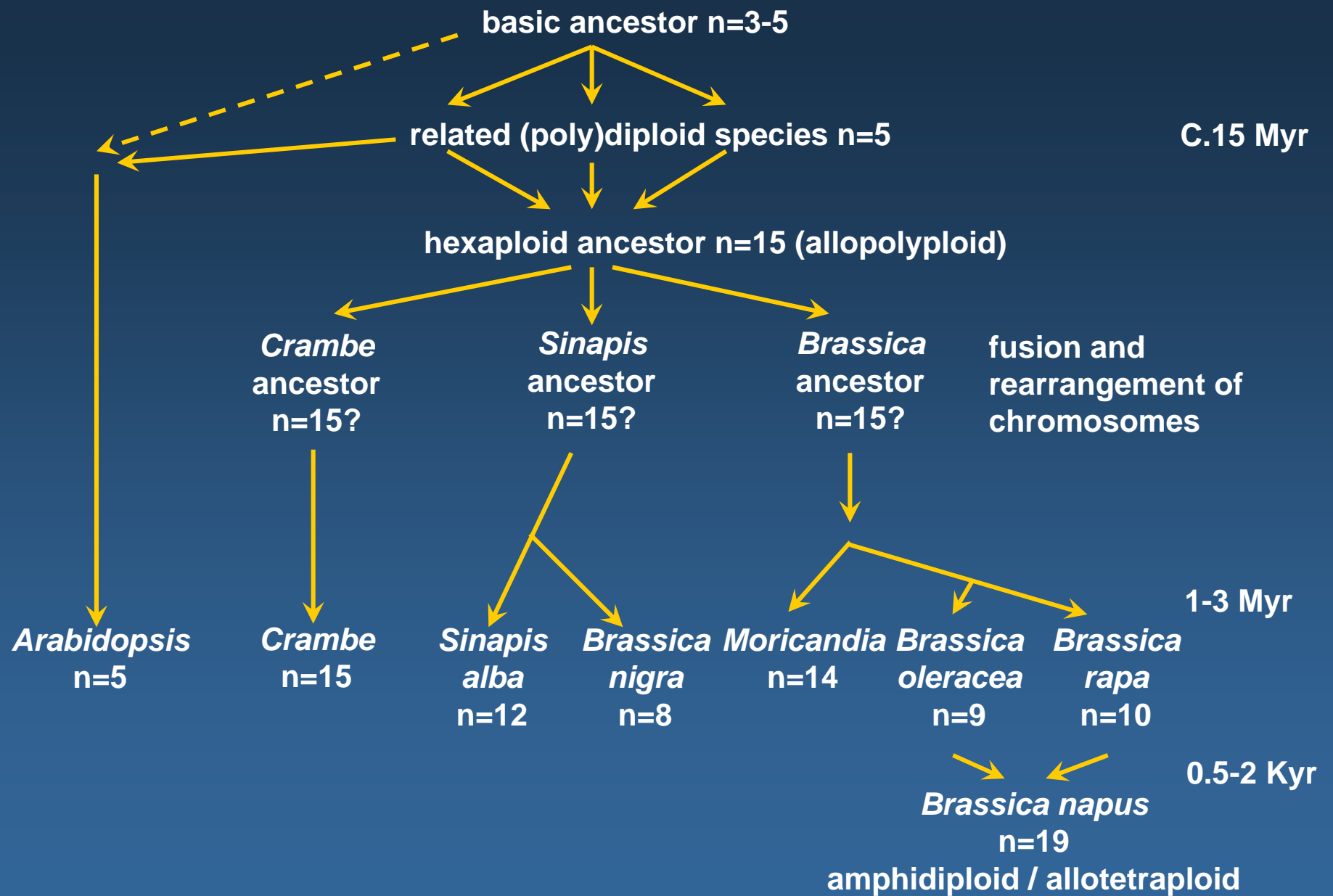


Model Crucifer
Arabidopsis thaliana ($2n=10$)
Genome “sequenced”

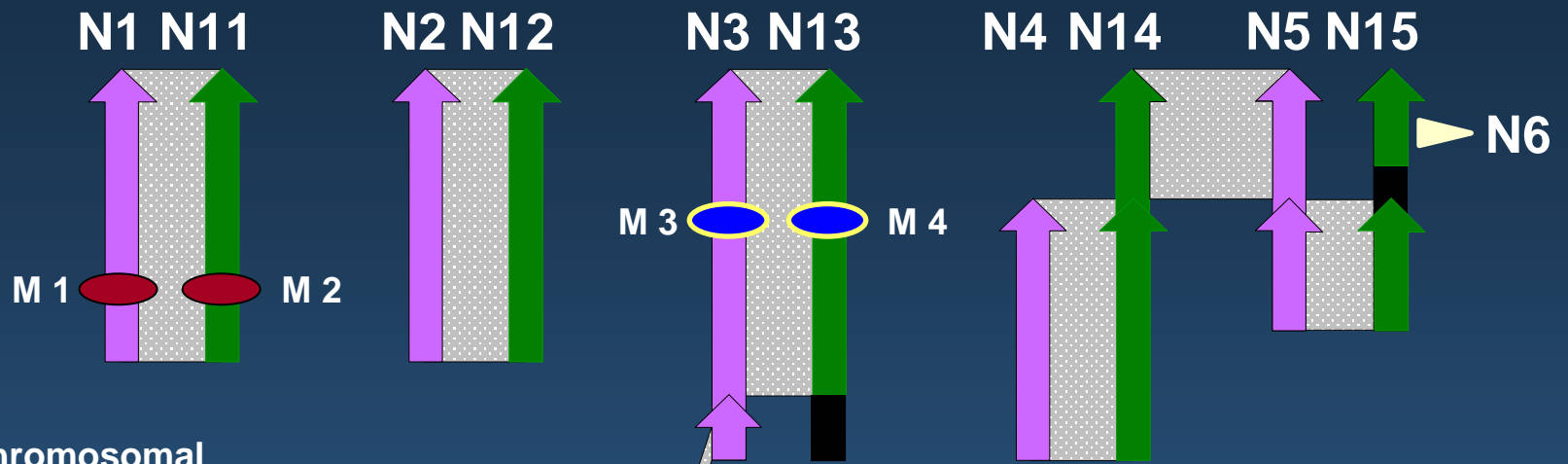


Agriculture and
Agri-Food Canada

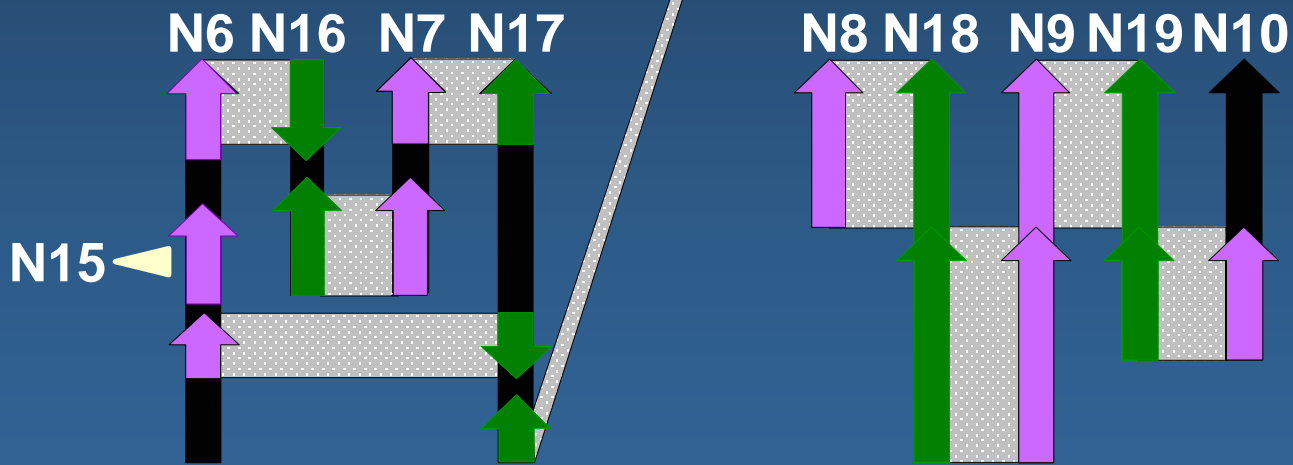
Agriculture et
Agroalimentaire Canada



The *B. napus* genome



16 major chromosomal rearrangements



Parkin et. al. 1995

N1-10 = A genome
N11-19 = C genome

M1 - M2 = 97% seq. similarity
M1 - M3 = 85% seq. similarity



Agriculture and
Agri-Food Canada

Agriculture et
Agroalimentaire Canada

Problems with an *Arabidopsis* array?

- Lower specificity and sensitivity
 - Due to divergence between *Brassica* and *Arabidopsis*
 - Useful for very conserved or highly expressed genes
- Unable to profile novel *Brassica* transcripts
- Unable to discriminate between *Brassica* gene family members
 - Or may only detect a subset of family members



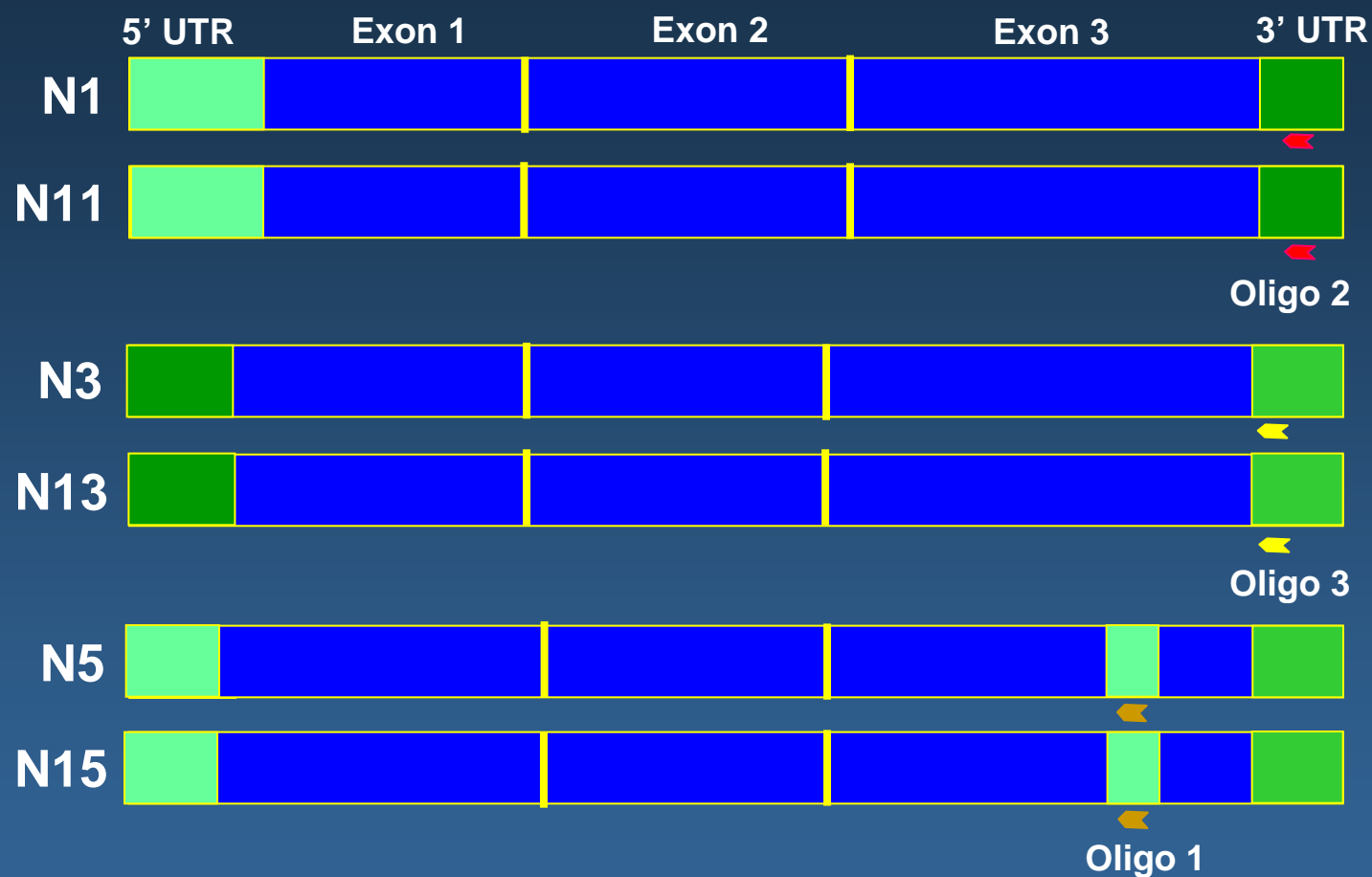
Brassica microarrays

The Strategy:

- Utilize >130K AAFC *B. napus* ESTs (cDNA - 5' and 3' ends)
- 3' ends and Untranslated Regions (UTRs)
 - Identify gene family variation
 - Design oligos representing genes
 - Resolve differential expression within families
- Useful for *Brassica* specific genes



Oligo design



Oligos specific to homoeologous gene pairs



Oligo design - 3'UTR

50-gil10010 BL3510F	CCAACCAA:CTA:AACT::C:TGTCTTTT:::G:C::TTC:TGTAATAGAAATTC:TAATATG
60-gil10010 BL3510F	CGTTTGTATCCAACCAA:CTA:AACT::C:TGTCTTTT:::G:C::TTC:TGTAATAGAAATTC:TAATATGT
70-gil10010 BL3510F	
gil10023 A5g08540_C1	
gil10024 A5g08540_C3	
BL1442F	TTAATCACAA:ACGTTTGTATCCAACCAA:CTA:AACT::C:TGTCTTTT:::G:C::TTC:TGTAATAGAAATTC:TAATATGTAGAACA
BL1434F	TTAATCACAA:ACGTTTGTATCCAACCAA:CTA:AACT::C:TGTCTTTT:::G:C::TTC:TGTAATAGAAATTC:TAATATGTAGAACAACAAATACAA
BL3681F	TTAATCACAA:ACGTTTGTATCCAACCAA:CTA:AACT::C:TGTCTTTT:::G:C::TTC:TGTAATAGAAATTC:TAATATGTAGAACAACAATACAAAGTTGTGGCTAATATAA
BL3632F	TTAATCACAA:ACGTTTGTATCCAACCAA:CTA:AACT::C:TGTCTTTT:::G:C::TTC:TGTAATAGAAATTC:TAATATGTAGAACAACAATACAAAGTTGTGGCTAATATAA
BL4967F	TTAATCACAA:ACGTTTGTATCCAACCAA:CTA:AACT::C:TGTCTTTT:::G:C::TTC:TGTAATAGAAATTC:TAATATGTAGAACAACAATACAAAGTTGTGGCTAATATAA
BL5234R	TTAATCACAA:ACGTTTGTATCCAACCAA:CTA:AACT::C:TGTCTTTT:::G:C::TTC:TGTAATAGAAATTC:TAATATGTAGAACAACAATACAAAGTTGTGGCTAATATAA
BL1934F	TTAATCACAA:ACGTTTGTATCCAACCAA:CTA:AACT::C:TGTCTTTT:::G:C::TTC:TGTAATAGAAATTC:TAATATGTAGAACAACAACAAAAAAAAAAAAAAAAAAAA
BL759R2	TTAATCACAA:ACGTTTGTATCCAACCAA:CTA:AACT::C:TGTCTTTT:::G:C::TTC:TGTAATAGAAATTC:TAATATGTAGAACAACAACAAAAAAAAAAAAAAAAAAAA
BL3510F	TTAATCACAA:ACGTTTGTATCCAACCAA:CTA:AACT::C:TGTCTTTT:::G:C::TTC:TGTAATAGAAATTC:TAATATGTAAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAA
50-gil10023 BL676R2	CCAGCCAAACTAGAT:TGTCCTTTTGATATAAAGACAGTTTGTGTAATAGAA
60-gil10023 BL676R2	GACGTTATTATCCAGCCAAACTAGAT:TGTCCTTTTGATATAAAGACAGTTTGTGTAATAGAA
70-gil10023 BL676R2	GTATGTGTACGACGTTATTATCCAGCCAAACTAGAT:TGTCCTTTTGATATAAAGACAGTTTGTGTAATAGAA
gil10025 A5g08540_C2	ATAGAAAAATCTAAGATTCAGCACAAACAATACCAGTTGTGGCTTATATCA
BL3058F	GTATGTGTACGACGTTATTATCCAGCCAAACTAGAT:TGTCCTTTTGATATAAAGACAGTTTGTGTAATAGAAAAATCTAAGATTCAGCACAAACAATACCAGTTGTGGCTTATATCA
BL3058R	GTATGTGTACGACGTTATTATCCAGCCAAACTAGAT:TGTCCTTTTGATATAAAGACAGTTTGTGTAATAGAAAAATCTAAGATTCAGCACAAACAATACCAGTTGTGGCTTATATC
BL2850F	GTATGTGTACGACGTTATTATCCAGCCAAACTAGAT:TGTCCTTTTGATATAAAGACAGTTTGTGTAATAGAAAAATCTAAGATTCAGCACAAACAATACCAGTTGTGGCTTATATCA
BL5498R	GTATGTGTACGACGTTATTATCCAGCCAAACTAGAT:TGTCCTTTTGATATAAAGACAGTTTGTGTAATAGAAAAATCTAAGATTCAGCACAAACAATACCAGTTGTGGCTTATATCA
BL5498R	GTATGTGTACGACGTTATTATCCAGCCAAACTAGAT:TGTCCTTTTGATATAAAGACAGTTTGTGTAATAGAAAAATCTAAGATTCAGCACAAACAATACCAGTTGTGGCTTATATCA
BL5553R	GTATGTGTACGACGTTATTATCCAGCCAAACTAGAT:TGTCCTTTTGATATAAAGACAGTTTGTGTAATAGAAAAATCTAAGATTCAGCACAAACAATACCAGTTGTGGCTTATATCA
BL2749F	GTATGTGTACGACGTTATTATCCAGCCAAACTAGAT:TGTCCTTTTGATATAAAGACAGTTTGTGTAATAGAAAAATCTAAGATTCAGCACAAACAATACCAGTTGTGGCTTATATCA

640 650 660 670 680 690 700 710 720 730 740

TTAATCACAAAGACGTTTGTATCCAACCAA:CTA:AACT::C:TGTCTTTT:::G:C::TTC:TGTAATAGAAATTC:TAATATGTAGAACAACAATACCAGTTGTGGCTTATATCA

- Design oligos to distinguish different members of gene families



Trial oligo set

Trial oligos

- 32 members of gene families from Bud and Leaf libraries (relatively even number of sequences for oligo design, easy to obtain tissue for RNA extraction, likely to detect differential expression)
- Cluster at 75% sequence similarity, pick representative sequence
- 50mer, 60mer and 70mer oligos designed to test for specificity and sensitivity (for a total of 96)
- Design oligos with OligoPicker

(Wang X., and B. Seed (2003): Bioinformatics 19:796-802)

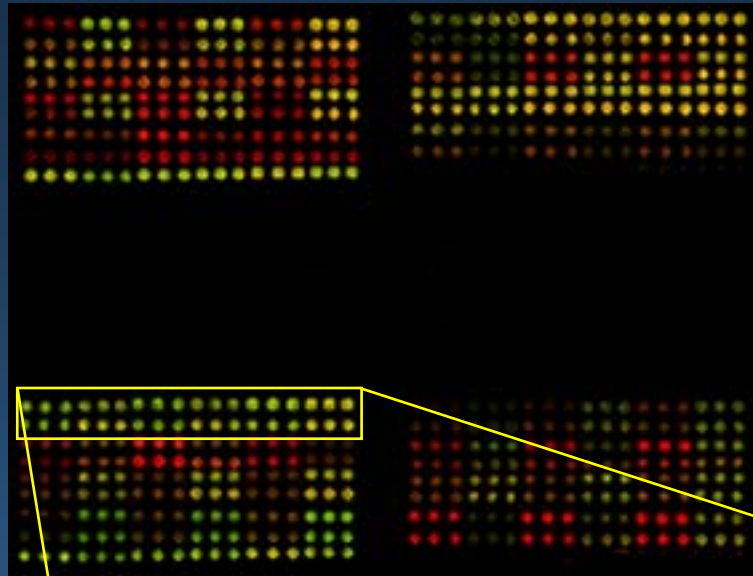
<http://pga.mgh.harvard.edu/oligopicker/index.html>



Agriculture and
Agri-Food Canada

Agriculture et
Agroalimentaire Canada

Trial oligo set

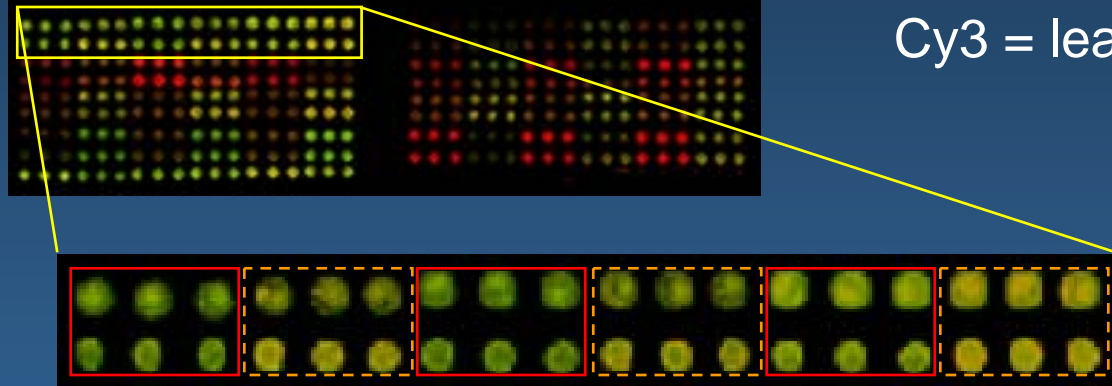


Example:

Super Amine slide (Telechem)

35% - 40% relative humidity

Cy3 = leaf mRNA, Cy5 = bud mRNA



Telechem printing buffer

Full Moon printing buffer

gene A - 50mer

60mer

70mer

gene B - 50mer

60mer

70mer



Agriculture and
Agri-Food Canada

Agriculture et
Agroalimentaire Canada

Global *Brassica* oligo design

- Two phase automated bioinformatics analysis:
- Phase 1
 - Focus on sequences with homology to *Arabidopsis* at the nucleotide level and identify *Brassica* gene family members
 - Should produce oligos with better sensitivity and specificity than corresponding *Arabidopsis* oligo
- Phase 2
 - Focus on sequences without homology to *Arabidopsis* at the nucleotide level but have homology at protein level to *Arabidopsis* or GenBank database representing other species
 - Provide *Brassica* specific oligos



Final set

- 18,000 *Brassica* 50mer oligos designed:
 - 16,000 unique, 2,000 cross-hybridizing
- Oligos screened based on annotated function:
 - Removed sequences associated with ribosome components, nucleotide metabolism and cellular respiration
- 15,000 sequences sent for synthesis.



Acknowledgments

Thanks to:

Chris Lewis

Branimir Gjetvaj

Derek Lydiate

Funding provided by:

UK NERC (Mike Kearsey, John Walsh, Ian Cooper)

AAFC Crop Genomics Initiative

Oligo synthesis by Illumina

Arrays will be available end of 2004 Genome Viewer - www.brassica.ca



Agriculture and
Agri-Food Canada

Agriculture et
Agroalimentaire Canada