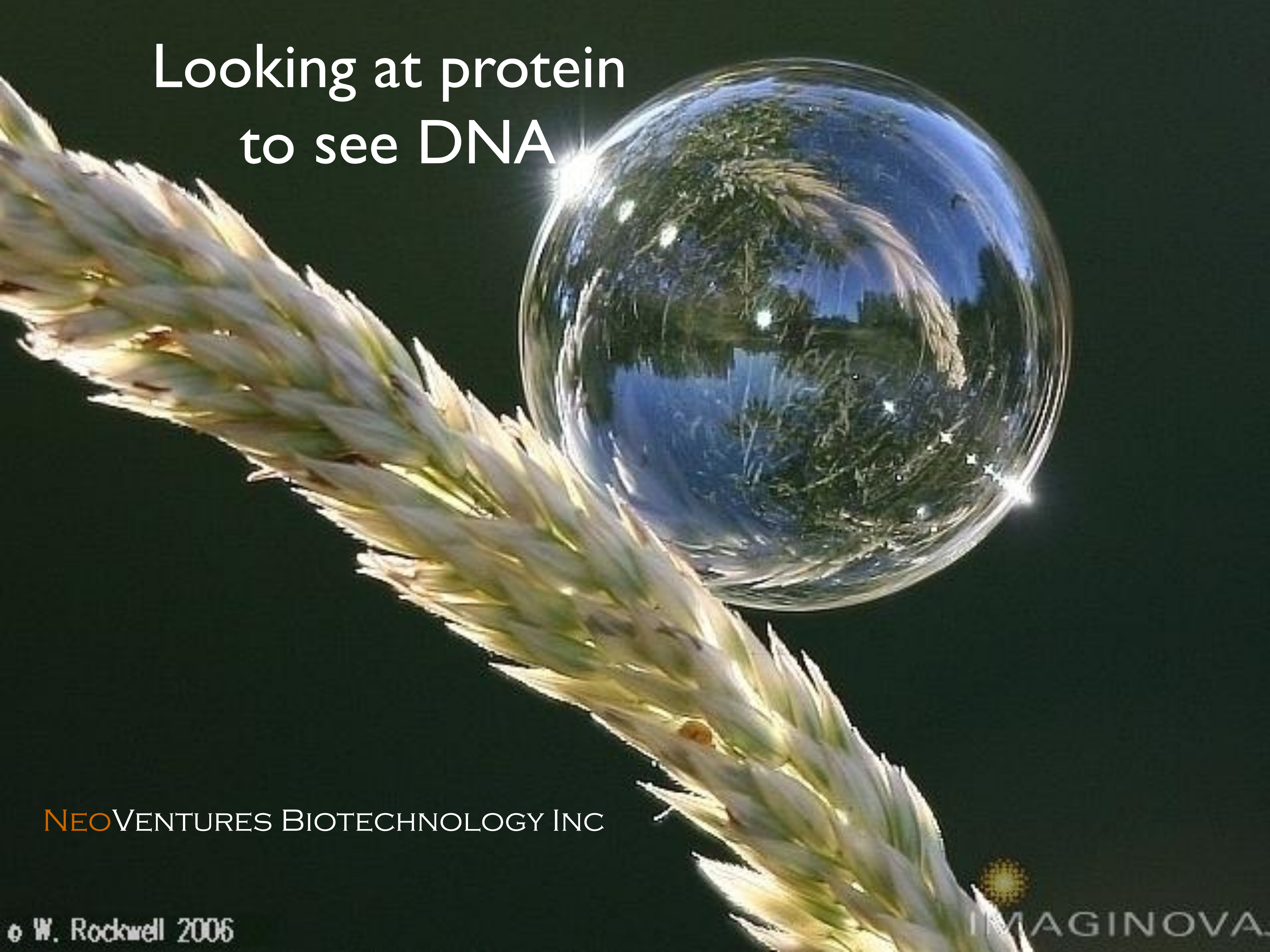
Looking at protein to see DNA

NeoVentures Biotechnology Inc



© W. Rockwell 2006

NEOVENTURES BIOTECHNOLOGY INC.

- Products
  - DNA ligand based mycotoxin testing
    - Single stranded DNA sequence selection
    - Mimic antibodies
  - Ochratoxin A strip test
    - 2 ppb sensitivity
    - Direct analysis of wine, beer and grain extracts
  - DNA ligands identified for zearalenone and aflatoxin B1

NeoVentures Biotechnology Inc.

- Variety Identification
  - Identification of proteins on intact grain
  - Determination of protein sequence variation among genotypes
  - Characterization of diagnostic peptides

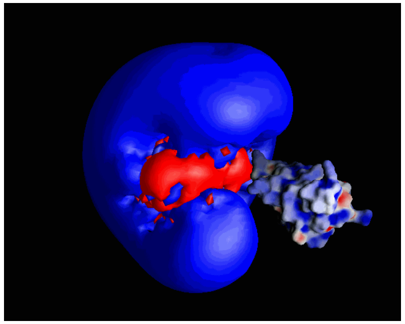
- Rapid, point of delivery, grain identification platform.

Product concept

Wash grain with extraction buffer

Capture target proteins in affinity column

Trypsin digest



Elute peptides onto detection strip

Determine genetic identity with bar code reader

Would this platform be useful for plant breeding?

Proteins vs. DNA

- Second generation DNA sequencing looks pretty awesome.
- Protein analysis is time consuming and expensive right?
  - Maybe not…

Protein analysis may yield more information/$ than DNA sequencing.

Tritin protein sequence

MAKNVDKPLFTATFNVQASSADYVTFINGIRNKLR
NPGHSSHNRPVLPPIEPNVPPSRWFHIVLKTSPASTGLTLATR
ADNLYWEGFKSSDGTWWELTPGLIPGATHVGFGGTYR
DLLGDTDKLTNVALGRQQMADAVTALYGRTKADKTSGPK
QQQAREAVTTLLLMVHEATRFQTVSGFVAGVLHPKEK
KSGKIGNEMKAQVNGWQDLSEALLKTDAKPPPGKAPAK

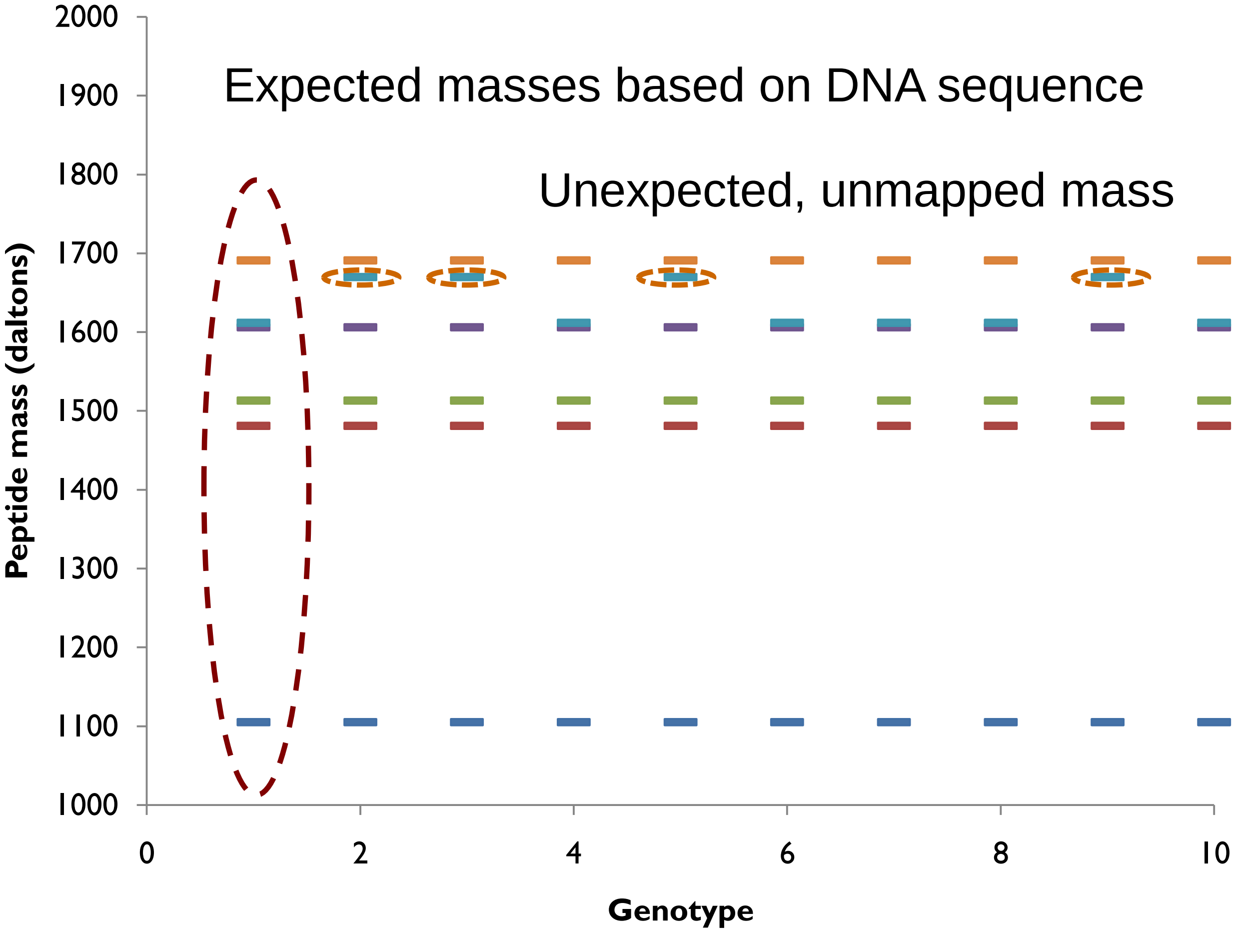
FTPIEKMGVRTAEQAAATLGILLFVQVPGGMTVAQALELFHKSGGK

Tritin is a ribosome inactivating protein found on the grain surface.

Peptides expected after trypsin digestion.

Sequence	Mass (Daltons)
NVDKPLFTATFNVQASSADYVTFINGIR	3088
NPGHSSHNRPVLPPIEPNVPPSR	2498
WFHIVLK	942
TSPASTGLTLATR	1275
ADNLYWEGFK	1242
SSDGTWWELTPGLIPGATHVGFGGTYR	2862
DLLGDTDK	876
LTNVALGR	843
QQMADAVTALYGR	1423
QQQAR	630
EAVTTLLLMVHEATR	1683
FQTVSGFVAGVLHPK	1586
IGNEMK	691
AQVNGWQDLSEALLK	1671
TDAKPPPGK	910
FTPIEK	734
TAEQAAATLGILLFVQVPGGMTVAQALELFHK	3324

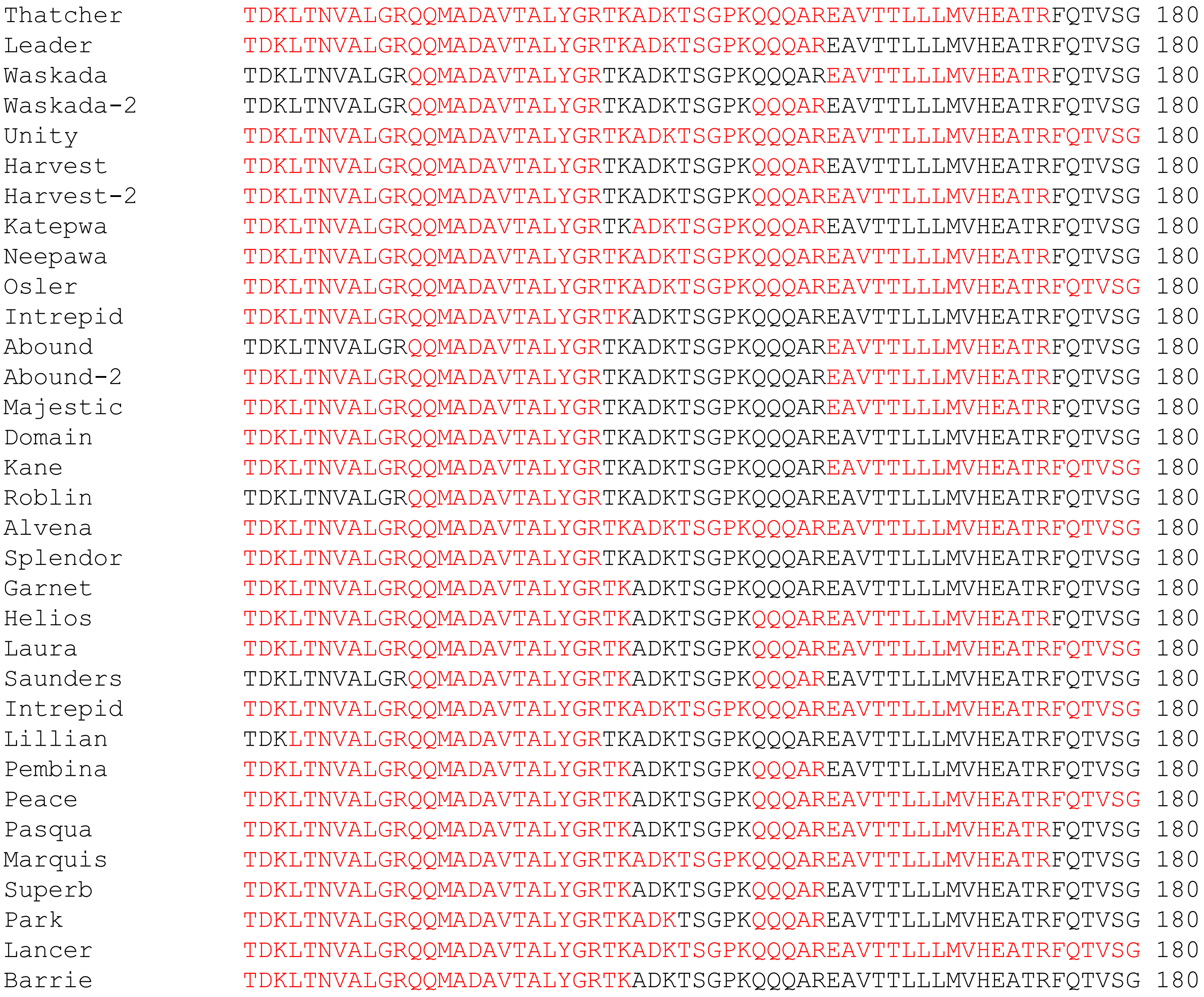
Mass spectra



Expected masses based on DNA sequence

Unexpected, unmapped mass

Mapping MALDI TOF peptides.



```
Thatcher        TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Leader          TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Waskada         TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Waskada-2       TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Unity           TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Harvest         TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Harvest-2       TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Katepwa         TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Neepawa         TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Osler           TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Intrepid        TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Abound          TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Abound-2        TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Majestic        TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Domain          TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Kane            TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Roblin          TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Alvena          TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Splendor        TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Garnet          TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Helios          TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Laura           TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Saunders        TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Intrepid        TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Lillian         TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Pembina         TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Peace           TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Pasqua          TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Marquis         TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Superb          TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Park            TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Lancer          TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
Barrie          TDKLTNVALGRQQMADAVTALYGRTKADKTSGPKQQQAREAVTTLLLMVHEATRFQTVSG 180
```

Predicting amino acid sequence

	Observed mass (Daltons)	Theoretical mass (Daltons)	Delta
EAVTTLLLMHEATR	1683.9345	1683.9098	0.0247
EAVTTLLLIHEATR	1665.9733	1665.9533	0.02
EAVTTLLLLHEATR	1665.9733	1665.9533	0.02

This represents possible sequences hits for an unmapped peptide mass.

Using the codons

gaggcggtgacgacgctgctcctc**atg**gtgcacgaggccacgcgg
E A V T T L L L M V H E A T R

Four possible SNPs

atc
ata
I

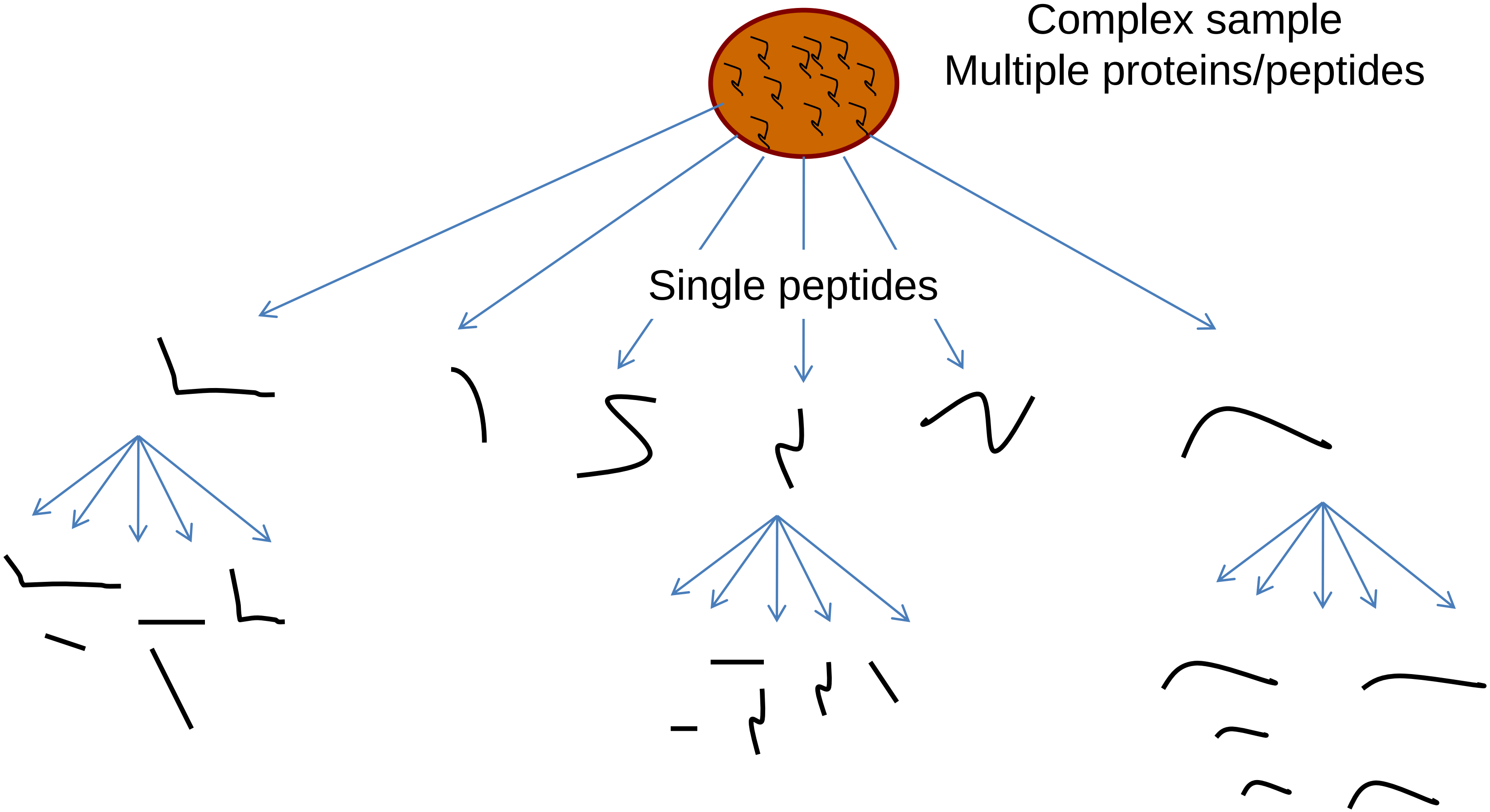
ttg
ctg
L

We have developed programs to automate this analysis.

Next steps

- Now, we could design four primers and confirm the sequence variation that way, but we have a better way.
- Peptide sequencing of a limited number of genotypes

MS/MS



Complex sample
Multiple proteins/peptides

Single peptides

Multiple fragments/peptide

MS/MS

- Generate a lot of peptide sequences
- Cost/analysis is an issue
- This is how we assign unknown masses.
    - Determine sequence of unmapped masses using a subset of genotypes.
    - Confirm sequence identities with directed sequencing
        - Identify clones with appropriate PCR primers.

– This is a high throughput, cost-effective system.

GPM Tornado (X! Hunter)

EAVTTLLLMVHEATR	1683.910	0.018
EAVTTLLLLVHEATR	1665.953	0.015

This was confirmed by sequence analysis.

- These peptides represent loci not alleles

Discovery of alleles

Tritin

SSDGTWWELTPGLIPGATHVGFGGTYR	2862.385
SSYGTWWELTPGLIPGATHVGFGGTYR	2910.4209

Allele present in Alsask, Domain, Imagine, Pembina, Waskada, Teal and Osler

Codon shift from GAC to TAC

In the same protein analysis

Glutathione-S-Transferase

GLPYEYAEEDLMAGK	Gst1a	1685.773
GLPYEYVEEDLMAGK	Gst1b	1713.804

IPVLLHDGRPVNESLIILQYLEDAFPDAPALLPSDPYAR	4328.297
AVNESLIILQYLEEAFPDAPALLPSDPYAR	3315.715
AVNESLIILQYLEEAFPHAPALLPSDPYAR	3337.747

P → A	CCC to GCC	(Locus)
D → H	GAC to CAC	(Allele)

Public Database www.neoventures.ca

We will be making all of our protein data publically available on our database.

- Locus sequences
- Allelic sequences
- Map positions
- SNPs
- 50 Canadian wheat varieties.

- Proteins targeted to date
  - Glutathione-S-Transferase
  - Polyphenol oxidase
  - Beta-amylase
  - Alpha-amylase inhibitors (mono/di/tetra)
  - Wheatwin
  - Catalase
  - Peroxidase

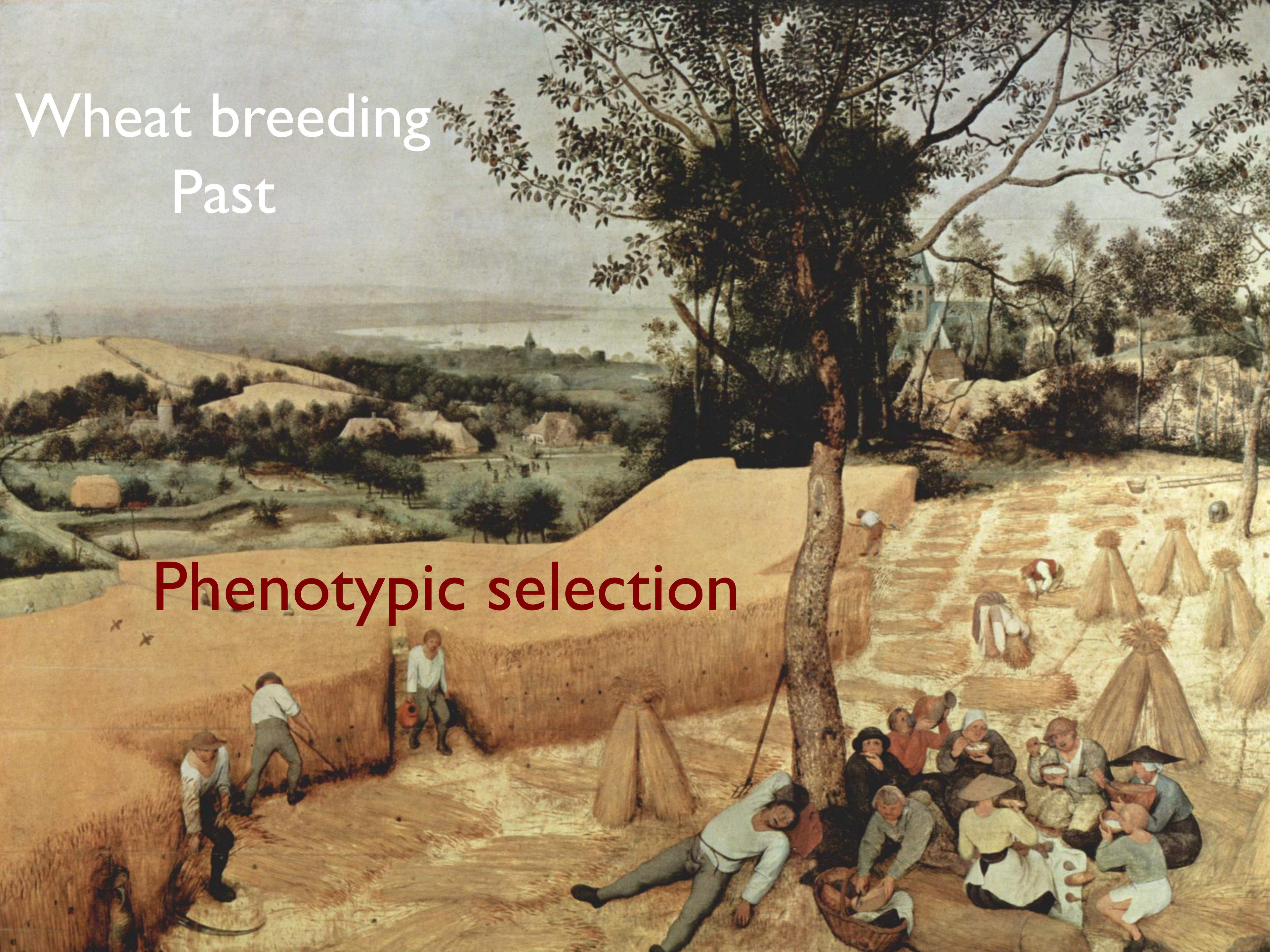
Summary

- Polyphenol oxidase
  - 3 Alleles at one locus
- Beta-amylase (ubiquitous)
  - 2 Alleles at one locus
- Catalase
  - 3 Alleles at one locus
  - 2 Alleles at another locus
- Tritin
  - Multiple alleles

Advantages

1. Multiple SNPs/locus
   - Great for haplotype analysis
2. SNPs in known proteins
   - Moving towards the next step in our understanding of variation in wheat

Wheat breeding
Past

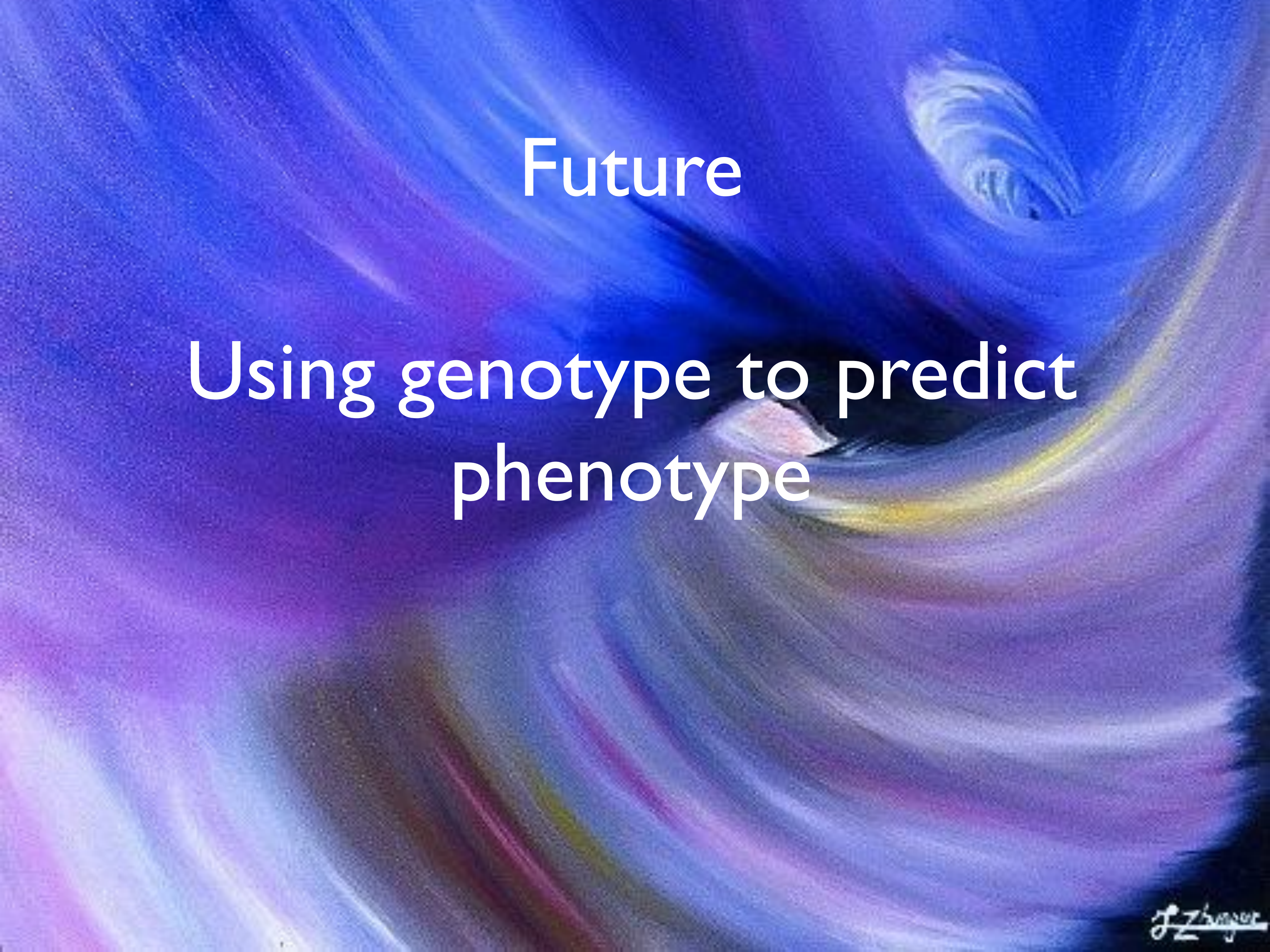


Phenotypic selection

DNA sequencing

Present

Correlation of phenotype to genotypic markers



Future

Using genotype to predict phenotype

The future

1. Identify sequence variation in proteins
2. Characterize variant sequence performance with recombinant proteins
3. Define desired recombination events
4. Drive improved genotypes based on genotypic understanding.
    - Crosses close to near-isogenic will work better.

NeoVentures Corporate Vision

- Provide SNP discovery service
- Provide rapid peptide analysis test kits
  - In-field tests of sequence variation
- Be global leaders in the application of proteomics to plant breeding

Partnership opportunities welcome;

www.neoventures.ca

gpenner@neoventures.ca

Thanks to our strategic partners and funding sponsors

- The Canadian Wheat Board
- Manitoba Rural Adaptation Council (AAFC)
- Industrial Research Assistance Program

The Neo Team

Gregory Penner	President & CEO
Ximena Vedoya	Vice President
Jenny Higham	Gene cloning
Marcie Chaudet	Peptide ligand selection
Theofanis Gkourasas	Protein analysis
Stephanie Santos	Gene mapping
Jorge Cruz	Senior research scientist
Linda Le	Mycotoxin ligand selection
Rachael Elston	Adminstrative support

Thank you