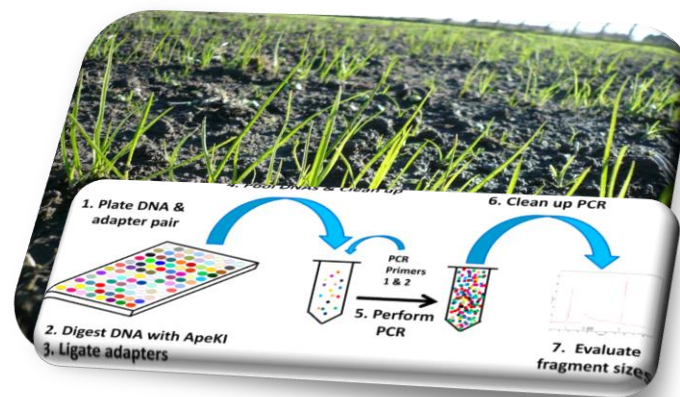# (toward) An Empirical Evaluation Of Genomic Selection In Perennial Ryegrass (*Lolium perenne* L.)

Marty Faville, Siva Ganesh, Roger Moraga, Zulfi Jahufer, Rob Elshire, Syd Easton, & Brent Barrett

NAAIC . TC . GBC
Lethbridge AB, Canada

marty.faville@agresearch.co.nz

1. Plate DNA & adapter pair
2. Digest DNA with ApeKI
3. Ligate adapters
5. Perform PCR
PCR Primers 1 & 2
6. Clean up PCR
7. Evaluate fragment sizes

Ministry of Business, Innovation & Employment

agresearch

Pastoral Genomics

GRASSLANDS innovation

DairyNz

BACKGROUND

Export value of pasture products to the NZ economy >$17.5 billion

Pasture feed base: perennial ryegrass/white clover mixed sward

Challenges for forages in pastures

- Meeting genetic potential of livestock ('feed gap')
- Intensification (herd size, stocking rate)
- Abiotic and biotic stresses
- Genotype x Site x Management = Complicated!

Improving the rate of genetic gain in forages is crucial for meeting the current and future needs of NZ farm systems

How can we best achieve this, and can markers & a
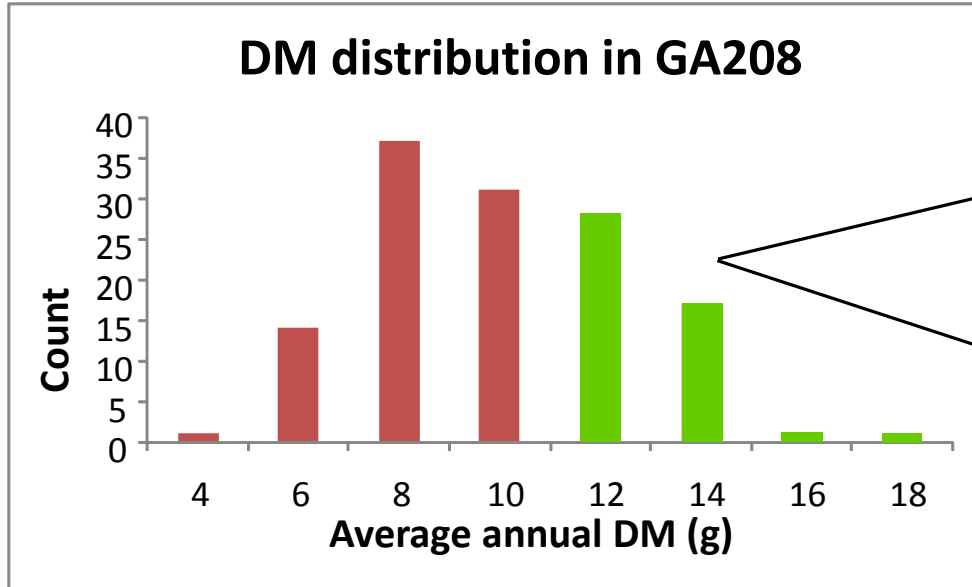*multi-disciplinary team help?*

agresearch

WHERE WE ARE AT

SINGLE MARKERS

BIG EFFECTS

ECONOMIC?

agresearch

Phenotype + Marker Selection for Ryegrass Yield
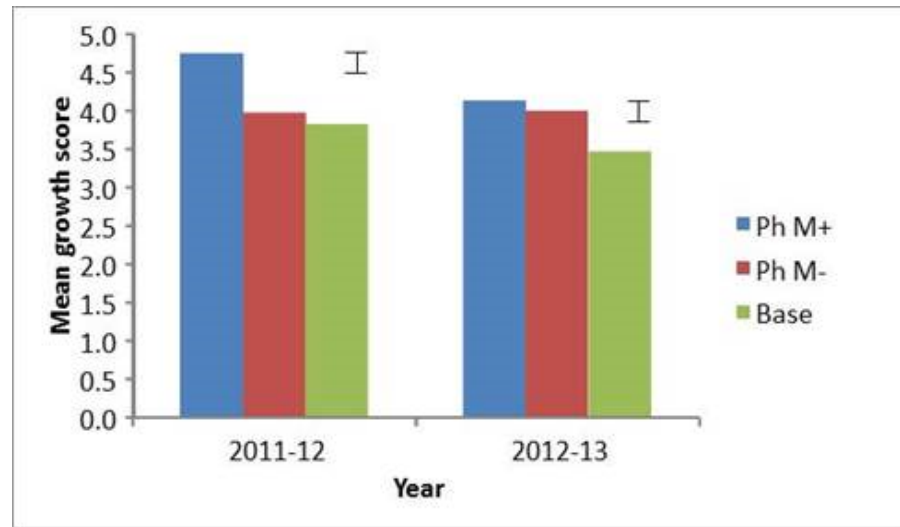
Phenotypically equivalent

DM distribution in GA208

Marker+ polycross (n=10)

12.8 g DM (±1.10)
3.46 GS (±0.32)

Marker- polycross (n=10)

12.2 g DM (±0.61)
3.44 GS (±0.22)

Count

Average annual DM (g)
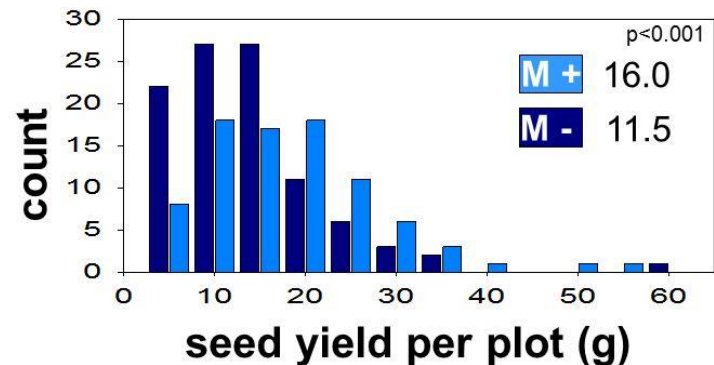
agresearch

Ryegrass DM Yield

Clover Persistence
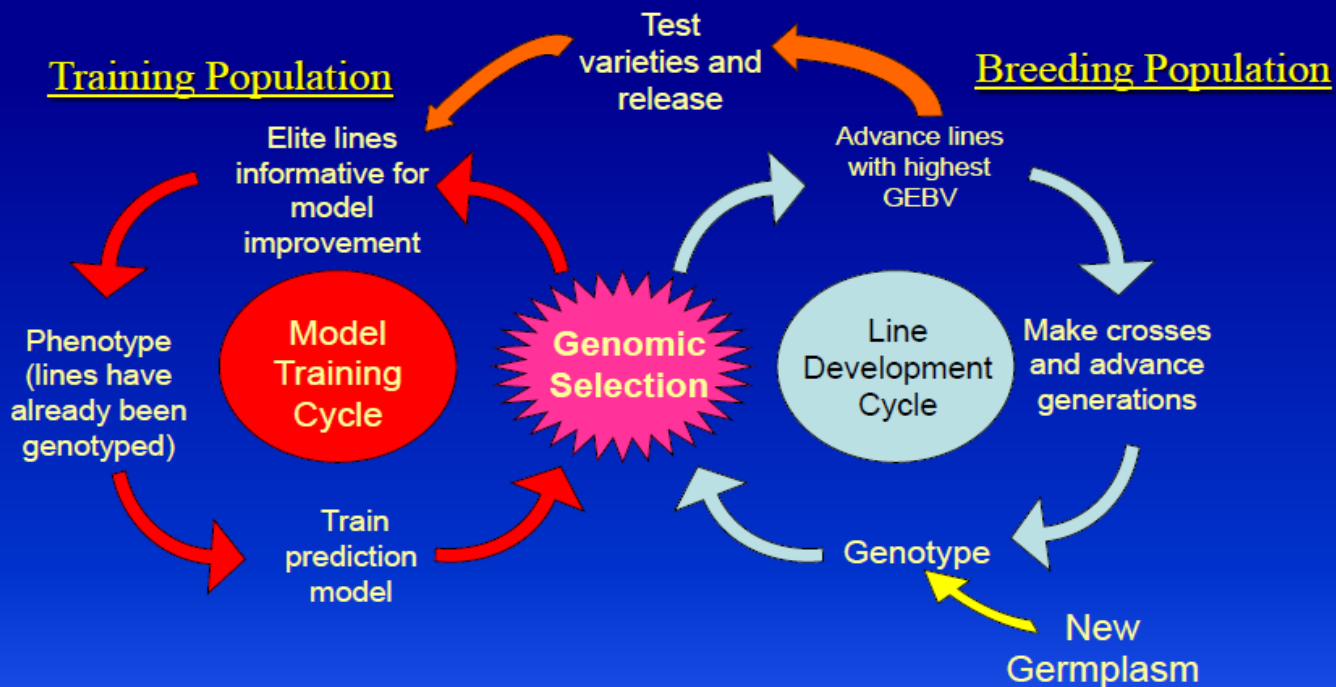
Poster 15
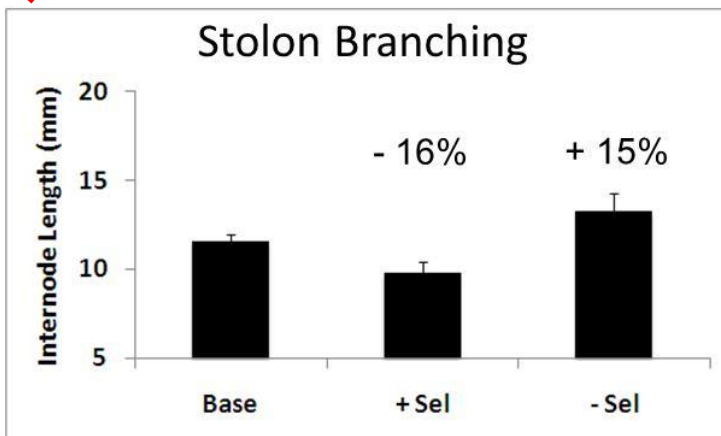
Clover Seed Yield

Barrett et al. (2009)
Proc. 5th MBFT

agresearch

WHERE WE ARE GOING

GS in a Plant Breeding Program

Heffner, Sorrells & Jannink. Crop Science 49:1-12, 2009

Genomic Selection (GS)

Dense markers predict genomic-estimated breeding values of selection candidates in breeding programmes

- Ideally capture all of the QTL that contribute to variation in a trait
- Work directly in breeding populations, no prior QTL discovery needed

Genotyping-by-Sequencing (GBS) makes GS an option for 'orphan' forage species

- No other accessible SNP resources (e.g. chips)
- Reference genomes only now becoming available
- It helps to have the expert in the team!

agresearch

Forage Value Index – Implications for Breeders!

New traits

Changing priorities

Heritabilities

Correlations

G x E & G x G interactions

Index Selection….

agresearch

Further Challenges

Phenotyping: Scale and Quality

Computational / Statistical

Trait Data Priority

GxGxSxM = Complicated

Scale to Australasian level: Relationships & Data

agresearch

Forage Breeding Model
Breeding Strategy & Operations to Optimise Gain per Unit Resource

FVI Economic Model

Bio-Physical Model (APSIM)

Forage Breeding Model

Define:
Value of Forage

Estimate:
FVI & Breeding Model in biophysical context

Calculate:
Genetic Gain in FVI per Unit Resource

agresearch
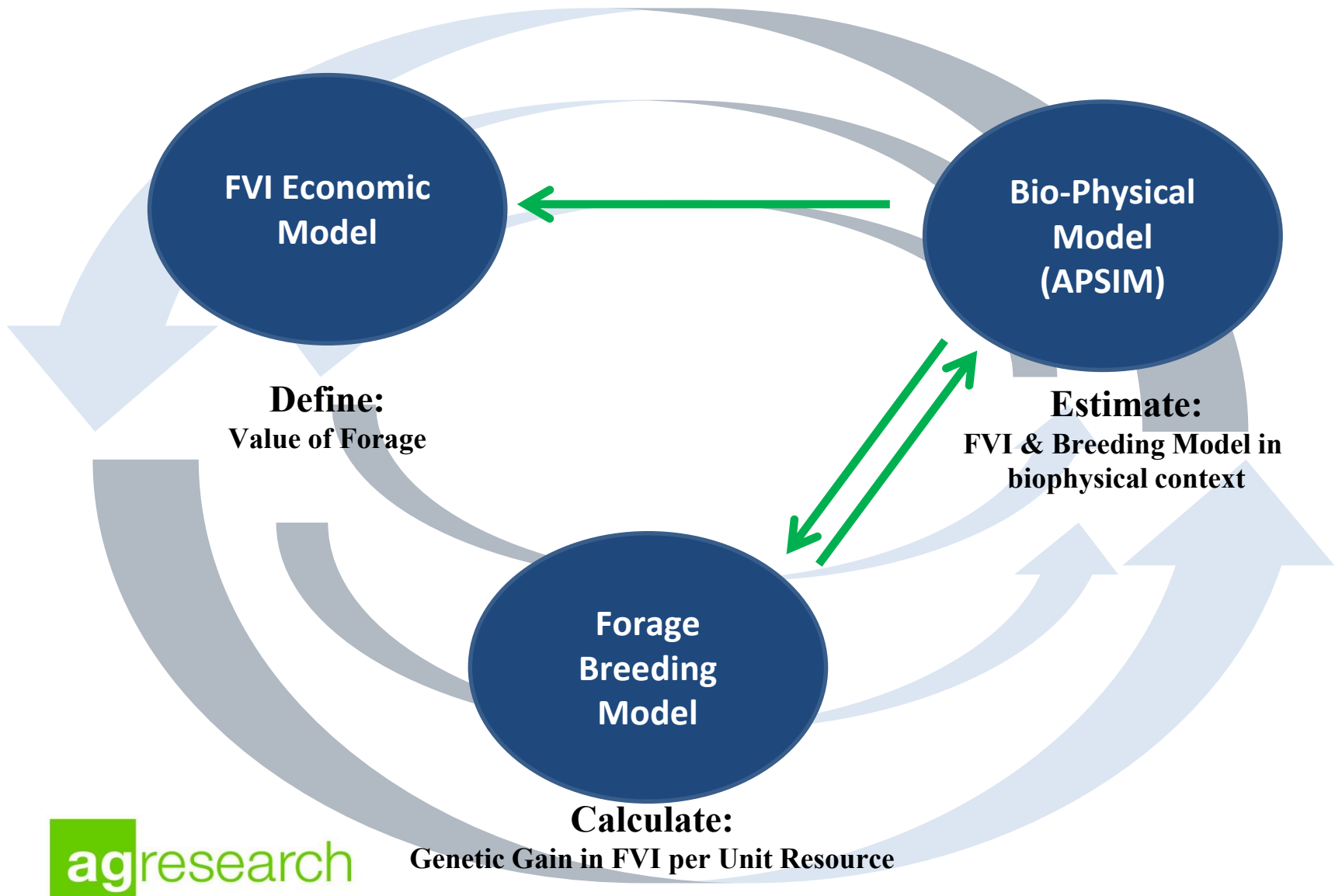
Empirical Work: Preliminary GEBVs In Ryegrass

▶ Populations GA194, PG1259 (n=211 x 4 reps in field, + DNA samples)

▶ Existing phenotypic datasets
- Single site, ~ two years seasonal data
- BLUPs for Vigour, DM production, Flowering Date

▶ Genotyping-by-sequencing (GBS)

⬇

c. 130K GBS tags

Filtering to SNPs

Calculate & Cross Validate GEBV's

Genotyping-by-Sequencing: SNP Calls & GEBVs

SNPs called using UNEAK (non-reference genome) pipeline (IGD)
- <u>3141</u> SNPs

SNPs re-called using TASSEL (reference genome pipeline)
- Full but fragmented reference genome sorted into 12 pseudo-molecules based on rice genome, used for mapping the GBS tags (University of Aärhus, Denmark)
- <u>13885</u> SNPs, 10624 mapped to a pseudo-molecule
- SNP subset for two pseudo-molecules = <u>2659</u> SNPs

Statistics - GEBVs
- 1 vs 2 stage models
- 7 Models x 9 Imputation methods
- 3 marker densities

Example Output: 1 v. 2 Stage x Low v. High SNPs

Models
Single Stage
RR- Ridge regression.
RK-Reproducing Kernel
RF-Random Forest

Markers (TASSEL)
Low = 2,659
High = 13,855

At least for small datasets,
statistical method matters!

agresearch

CROSS VALIDATION ACCURACY

Trait	CV accuracy, *r* (SE)		
	UNEAK pipeline (3K SNPs)	REF pipeline (10K SNPs)	REF pipeline (2K SNPs, C1 & 2)
Flowering time	0.55 (0.0041)	0.57 (0.0039)	0.53 (0.0043)
Vigour score (annual)	0.35 (0.0074)	0.31 (0.0062)	0.26 (0.0108)
Dry matter (annual)	0.34 (0.0054)	0.29 (0.0067)	0.22 (0.0104)

! More Markers ≠ Higher Accuracy

LD? Marker Data?   Family Structure?

agresearch

Ongoing Work

Assess relatedness/structure influence on GEBVs in this set

Further testing of statistical procedures and imputation methods

Reference genome-based SNP calling to obtain >20K SNPs
- Evaluation of marker density on GS prediction accuracy
- Mapping of GBS Tags, LD assessment
- Full reference genome (University of Aarhus, Denmark)

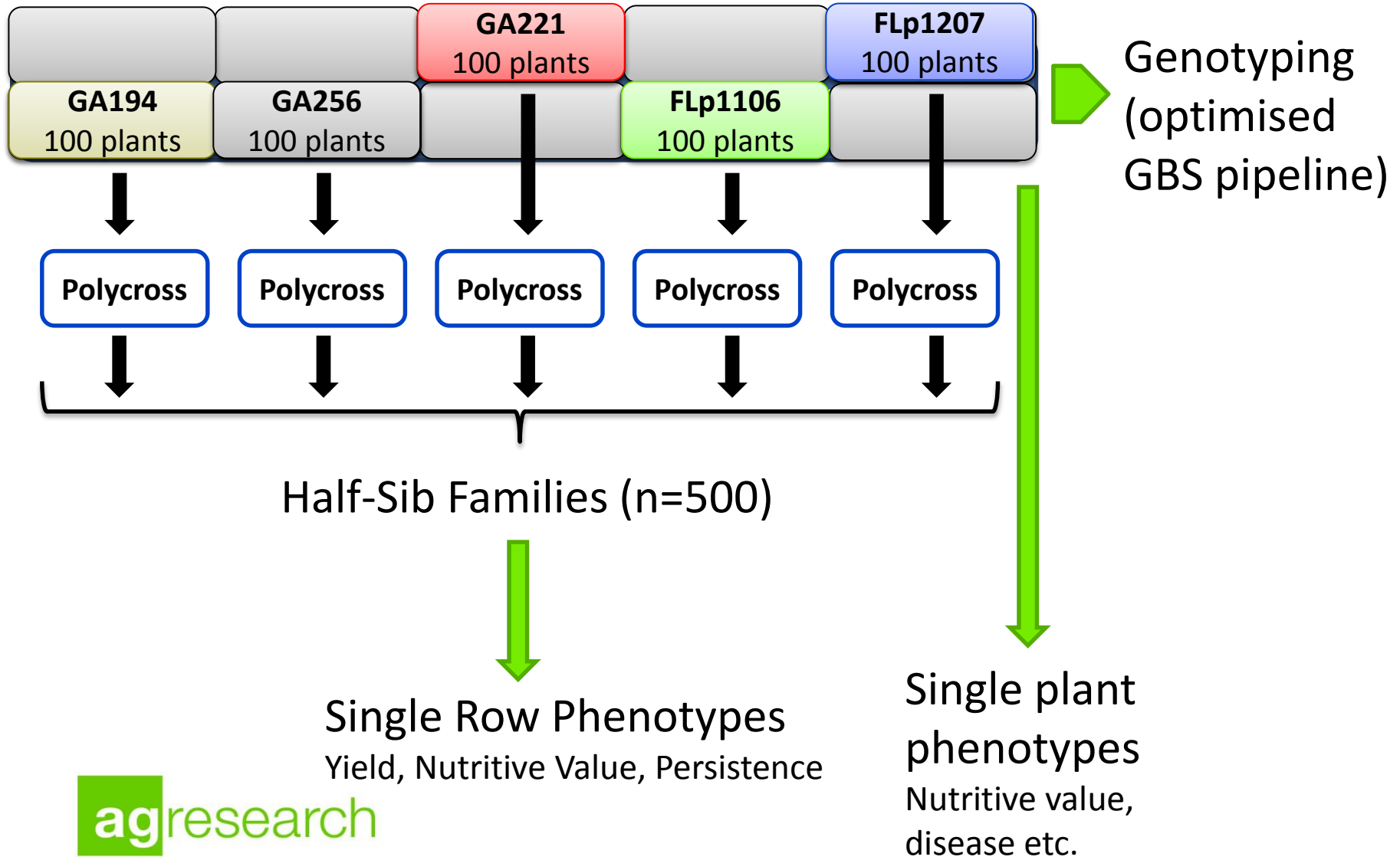Divergent selections from PG1259 and GA194 (low vs. high GEBV)
- Field evaluation of progeny commencing 2014-15

Additional training population & datasets from 2014-15
- Extension to wider breeding programmes

agresearch

Ryegrass Training Population Development

TRAINING POPULATION PHENOTYPING

Eight trials at five NZ locations established in 2013.

Half-Sib progeny rows (n>8000)

Emphasis on environmental replication over field replication

Kerikeri
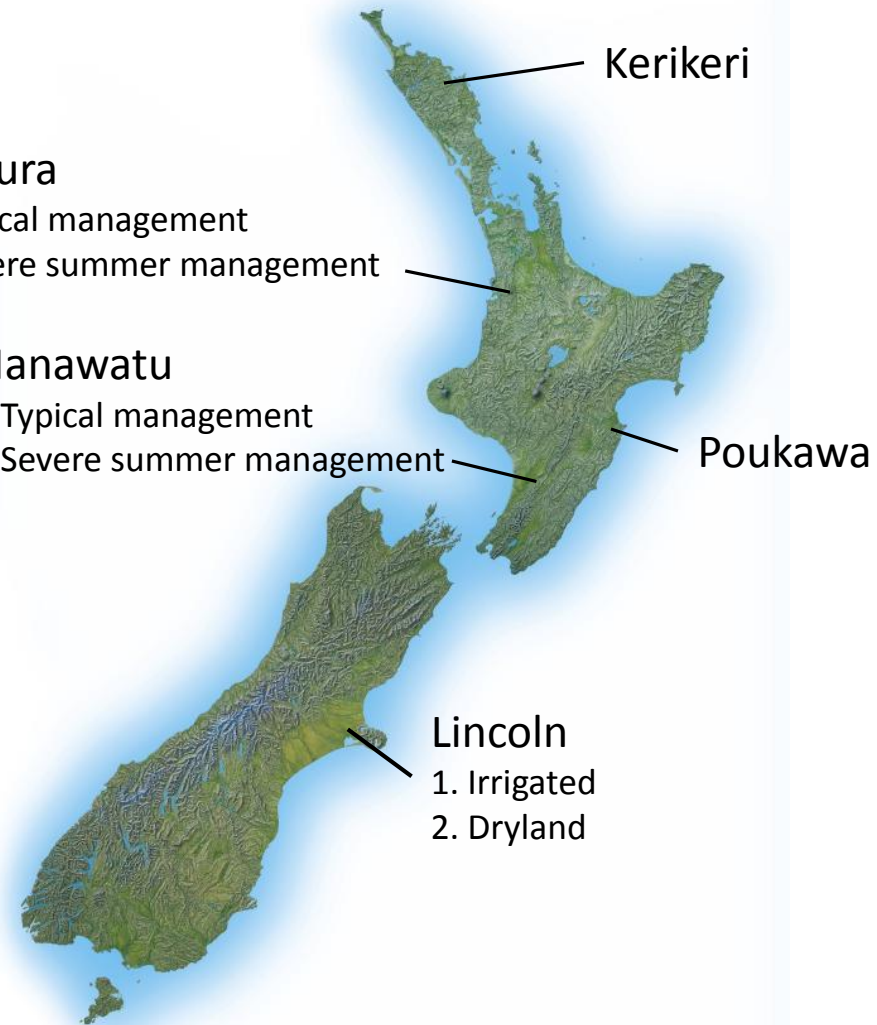
Ruakura
1. Typical management
2. Severe summer management

Manawatu
1. Typical management
2. Severe summer management

Poukawa

Lincoln
1. Irrigated
2. Dryland

FIRST FULL YEAR OF TRAINING POPULATION

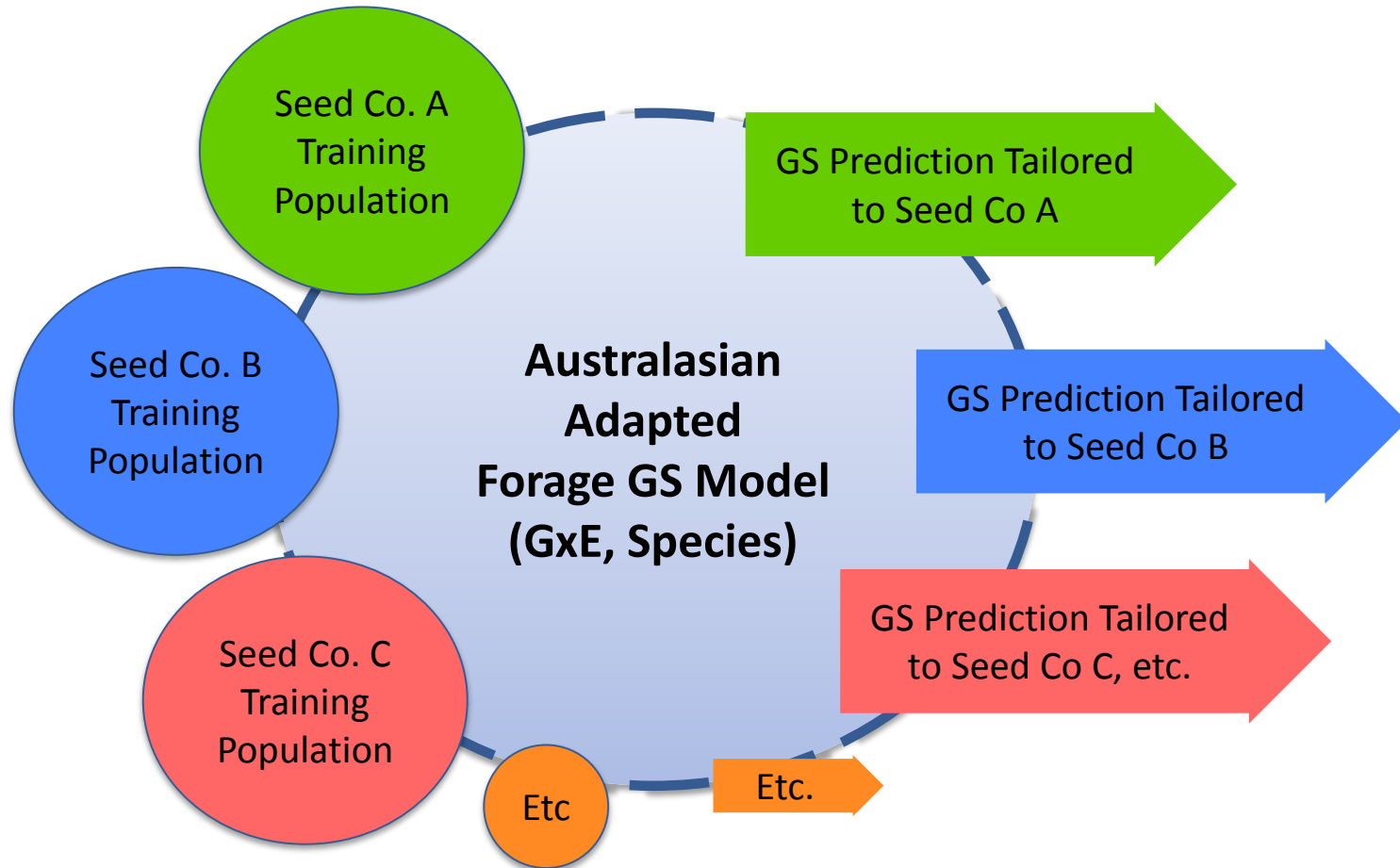8,000+ half sib rows is an awesome resource!

Phenotyping bottlenecks!!

Challenges to standardise management and measurement

Focus on seasonal yield & forage quality / composition

>2,500 HS progeny rows through wet lab fibre chemistry, soluble carbohydrates, field and lab bench NIRS completed autumn 2014

agresearch

Scaling Up (2) – A Pan Australian Approach

Seed Co. A Training Population

Seed Co. B Training Population

Seed Co. C Training Population

Etc

Australasian Adapted Forage GS Model (GxE, Species)

GS Prediction Tailored to Seed Co A

GS Prediction Tailored to Seed Co B

GS Prediction Tailored to Seed Co C, etc.

Etc.

agresearch

Generalising To An Australasian Model
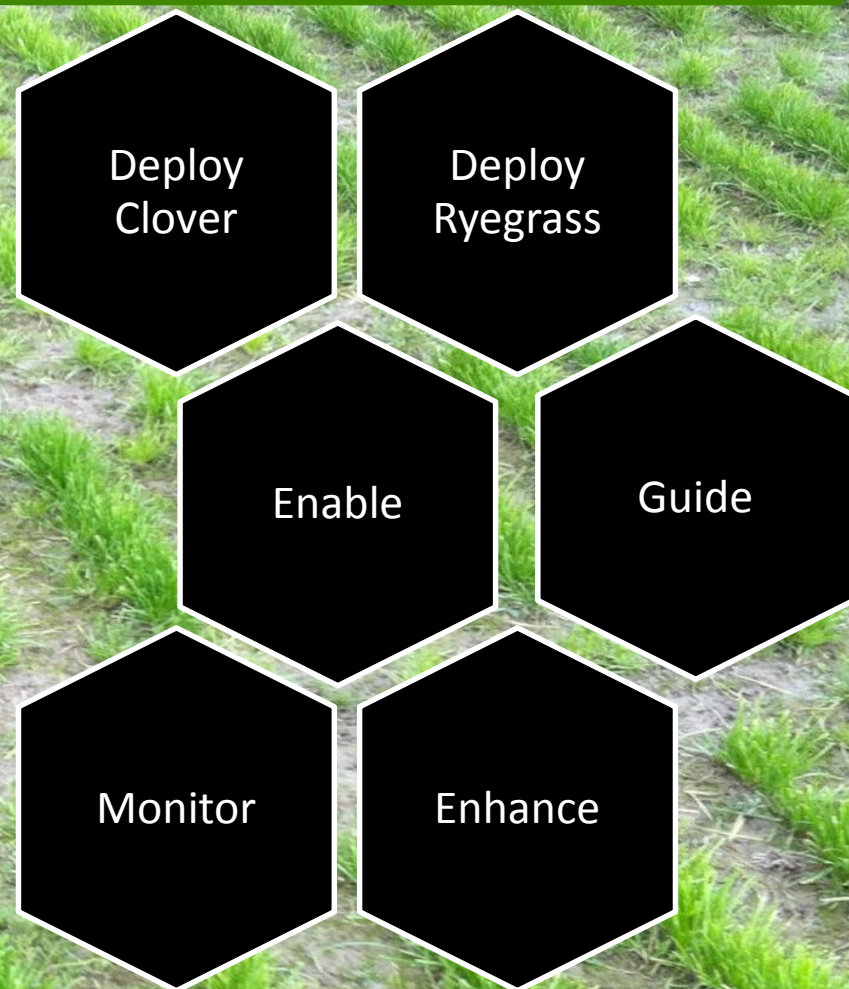
New Seed Firm Partnerships

Other Population Structures / Data Sources: individuals, families, bulks. Historic phenotypes?

Addressing the Phenotyping Bottleneck = technology + recruitment

Addressing the Computational Bottleneck = recruitment

Capability & Capacity: Field Breeding, Quantitative Genetics

agresearch

Genomic Selection: Proposed Workstreams

Final Comments

QTL-targeted MAS in forage breeding populations is workable, but not generally economic for key traits in ryegrass and white clover

Genomic selection is a logical progression from single/multi-marker MAS and may accelerate genetic gain for complex traits...*but it ain't easy!*

Upside: we're in the field and working in large trials of family structured material

Other species, traits, single gene markers, introgression all on the horizon

How will we know if it worked? Robust baseline monitoring.

agresearch

Acknowledgements

Tom Lyons
Benjamin Franzmayr
Jana Schmidt
Craig Anderson
Casey Flay
Tony Hilditch
Won Hong
Jessica O'Connor
Jason Trethewey

Mike Casler - USDA

Torben Asp