# Common Bean Coordinated Agricultural Project "BeanCAP"
# 2011 Advisory Board Meeting at PAG XIX



## Marker development:
## Perry Cregan, David Hyten & Qijian (Jian) Song

# Illumina GoldenGate
# Single Nucleotide Polymorphism (SNP) Analysis

## 1536 SNPs derived from a number of sources

- ◆ 1050 SNPs from a sequence analysis of Jalo EEP558 and BAT93 using "Next generation" sequence analysis (Roche 454 and Illumina Genome Analyzer) ★

- ◆ 486 SNPs from Sanger sequence analysis of common bean genes and PCR products derived from common bean using primers designed to soybean genes

- ◆ Analysis of common bean genotypes
  - ➤ 192 diverse genotypes representing the major market classes
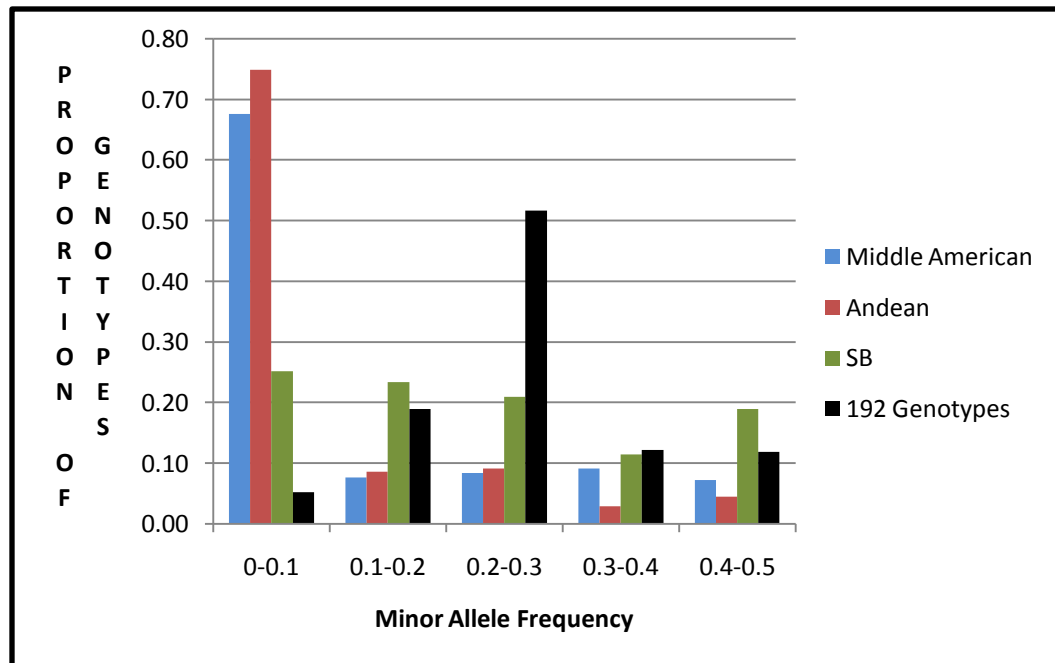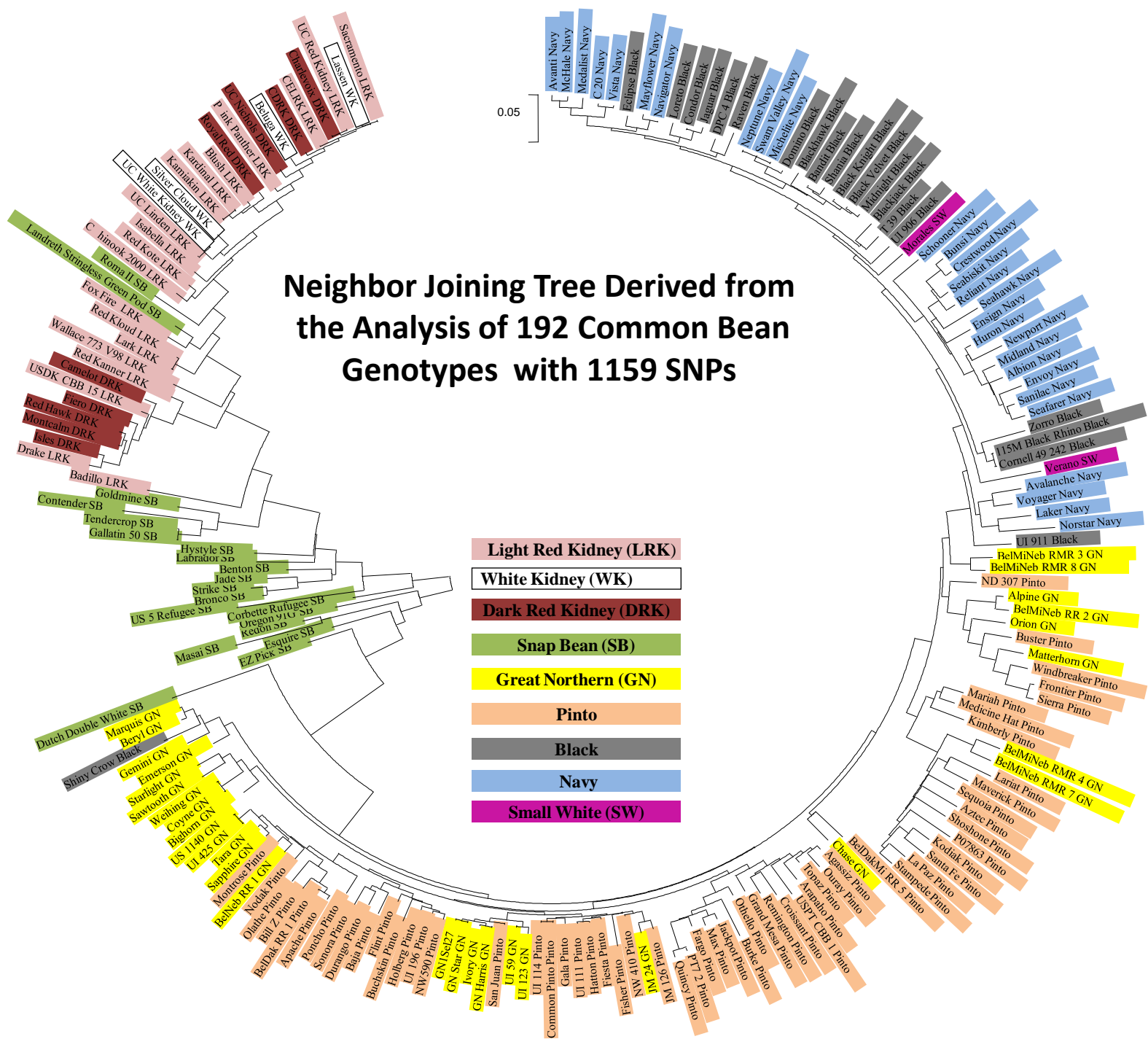  - ➤ 278 F2 lines of the Stampede x Red Hawk mapping population

★ For details see: Hyten, D.L., et al. 2010. High-throughput SNP discovery and assay development in Common Bean. BMC Genomics 11:475-483.

# Minor Allele Frequencies of 1159 SNPs Analyzed in 192 Common Bean Genotypes

| Minor Allele Frequency | Meso-american genotypes | Andean genotypes | Snap Bean genotypes | 192 Genotypes | Meso-american genotypes | Andean genotypes | Snap Bean genotypes | 192 genotypes |
|---|---|---|---|---|---|---|---|---|
| 0-0.1 | 782 | 868 | 292 | 61 | 0.676 | 0.749 | 0.252 | 0.053 |
| 0.1-0.2 | 89 | 99 | 271 | 220 | 0.077 | 0.085 | 0.234 | 0.190 |
| 0.2-0.3 | 97 | 106 | 243 | 599 | 0.084 | 0.091 | 0.210 | 0.517 |
| 0.3-0.4 | 105 | 34 | 133 | 141 | 0.091 | 0.029 | 0.115 | 0.122 |
| 0.4-0.5 | 84 | 52 | 220 | 138 | 0.073 | 0.045 | 0.190 | 0.119 |
| TOTAL | 1157 | 1159 | 1159 | 1159 | 1 | 1 | 1 | 1 |

Neighbor Joining Tree Derived from the Analysis of 192 Common Bean Genotypes with 1159 SNPs

Legend:
- Light Red Kidney (LRK)
- White Kidney (WK)
- Dark Red Kidney (DRK)
- Snap Bean (SB)
- Great Northern (GN)
- Pinto
- Black
- Navy
- Small White (SW)
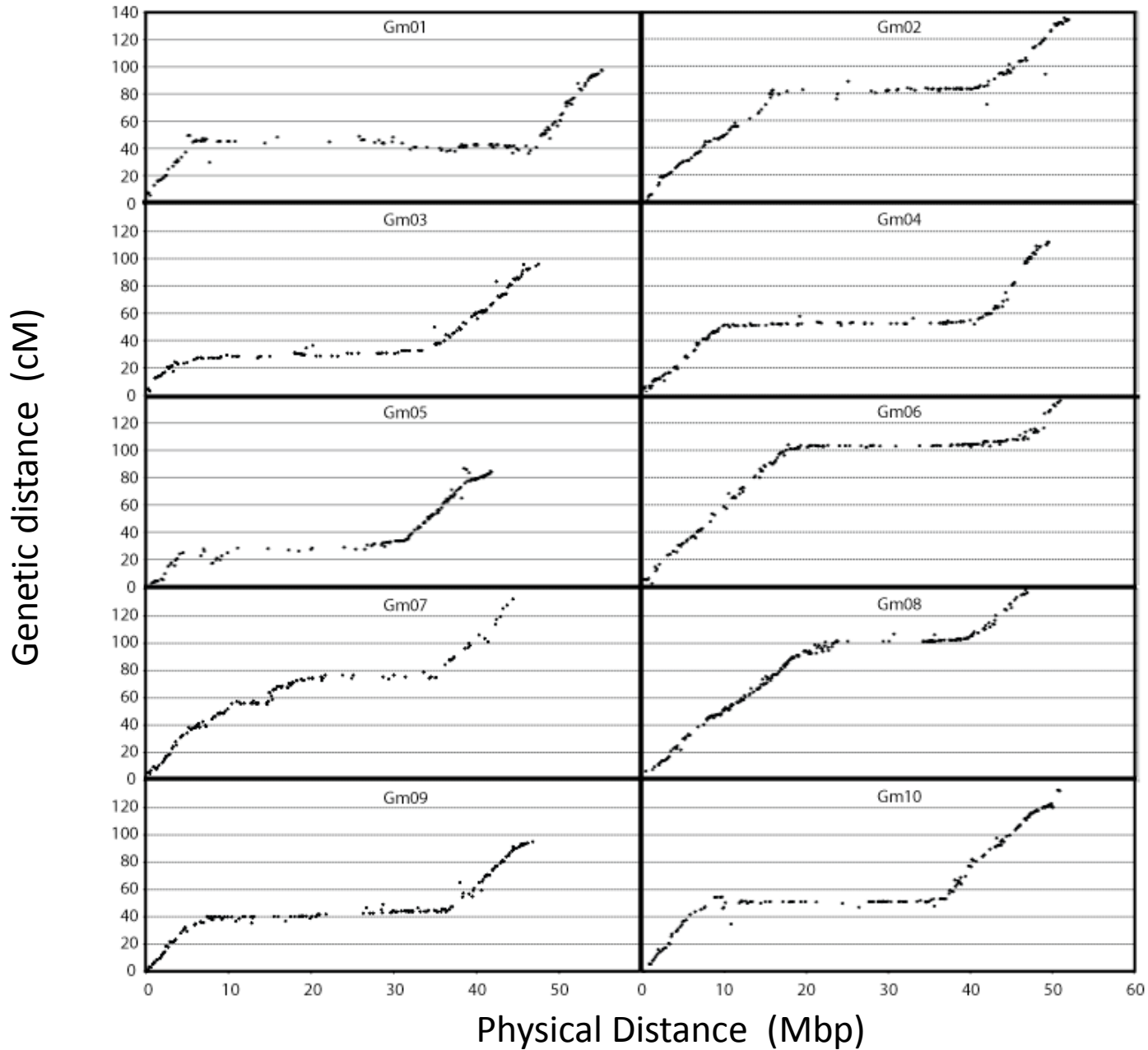
Stampede x Red Hawk Genetic Map (649 SNPs from PvOPA-1 + 25 NDSU Markers)
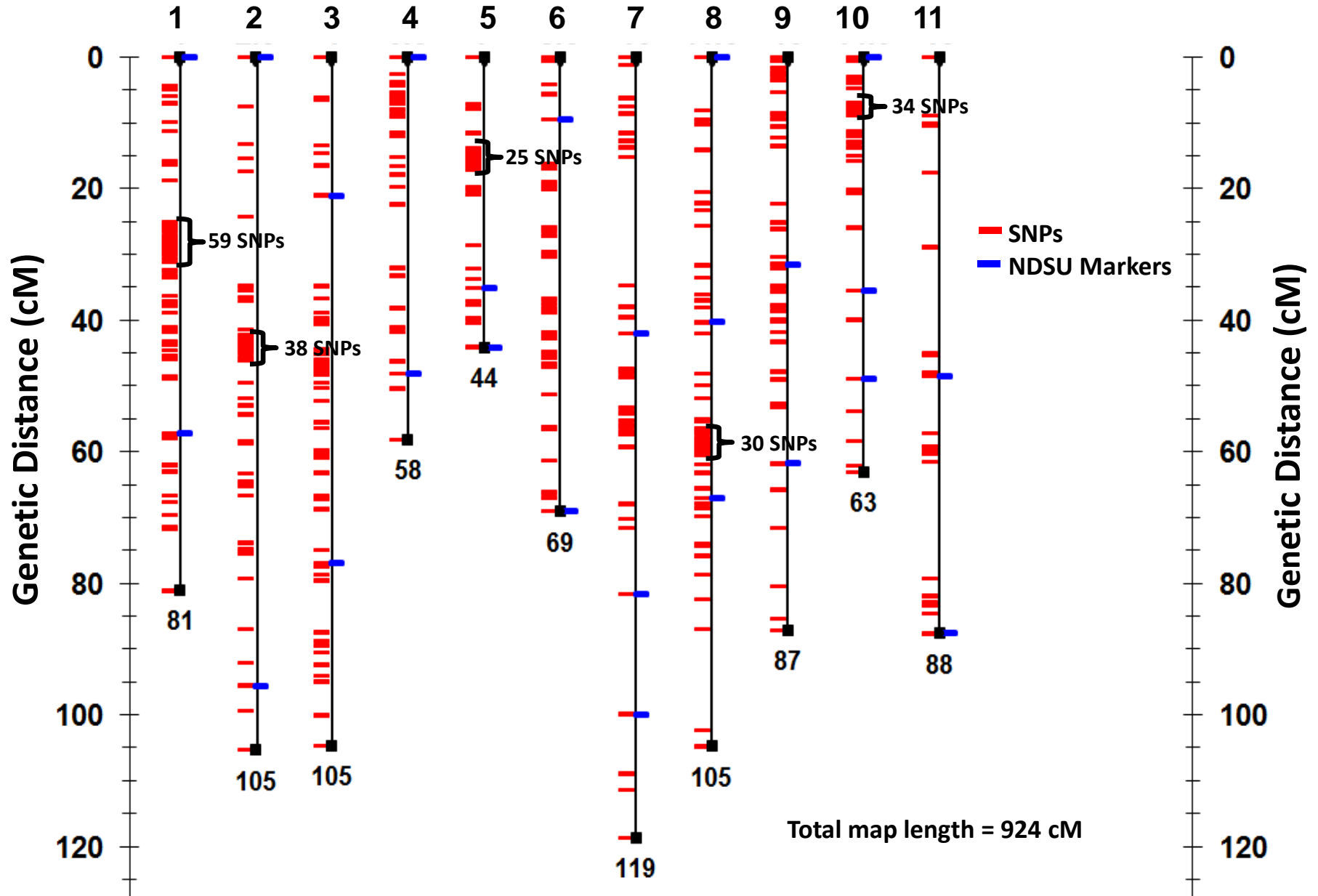
A Comparison of Genetic vs. Physical Distance in Soybean

# Development of a 6,000 SNP Illumina Infinium GeneChip

◆ Identify and select SNPs with high levels of polymorphism within market classes

➢ Conduct "next generation" whole genome DNA sequence analysis common bean genotypes

➢ Obtain whole genome sequence data from two or three genotypes per market class

◆ Position SNPs based upon genetic distance across the 11 Common Bean chromosomes

➢ Identify heterochromatic regions based upon the Stampede x Red Hawk map

➢ Position SNPs in a large number of the larger sequence contigs developed from the 14X whole genome sequence of G19833

# Illumina Genome Analyzer Whole Genome Sequence Obtained from 19 Bean Genotypes

| Cultivar Name | Market Class |
|---|---|
| Red Hawk | Dark Red Kidney |
| Fiero | Dark Red Kidney |
| California Early LRK | Light Red Kidney |
| Kardinal | Light Red Kidney |
| Lark | Light Red Kidney |
| UC_White_Kidney | White Kidney |
| Cornell_49_242 | Black |
| T-39 | Black |
| UI_906 | Black |
| Laker | Navy |
| C-20 | Navy |
| Michelite | Navy |
| Buckskin | Pinto |
| Stampede | Pinto |
| Sierra | Pinto |
| Gemini | Great Northern |
| Matterhorn | Great Northern |
| BAT93 | Bayo |
| Jalo EEP558 | Canário |

**No Snap Beans were included in the whole genome sequence analysis**

Neighbor Joining Tree Derived from the Analysis of 192 Common Bean Genotypes with 1159 SNPs

0.05

17 Cultivars for which Whole Genome Sequence data were obtained

Light Red Kidney (LRK)
White Kidney (WK)
Dark Red Kidney (DRK)
Snap Bean (SB)
Great Northern (GN)
Pinto
Black
Navy
Small White (SW)

# Preparation of DNA for Whole Genome Sequence Analysis Using the Illumina Genome Analyzer

➢ Genomic DNA was isolated from leaf tissue of the 19 common bean genotypes

➢ Genomic DNA was fragmented using a DNA fragmentase (New England Biolabs)

➢ Size selection was used to isolate fragments in the 250 to 400 bp size range.

➢ DNA was end repaired, an A nucleotide added to the 3' ends and indexed adaptors (barcodes) were ligated to the fragment ends.

➢ The DNA was PCR amplified to obtain sufficient DNA for analysis on the Illumina Genome Analyzer.

➢ Paired-end sequencing was used to obtain 115 bp of sequence from each end of the genomic fragments.

# Analysis of Whole Genome Sequence Data for SNP Discovery

- ➢ **The Illumina Off-Line Basecaller V1.8 software was used for base calling and CASAVA V1.7 was for demultiplexing.**

- ➢ **Approximately, 19.6 billion bases of DNA sequence data were obtained.**

- ➢ **The paired-end sequence reads from each genotype were aligned to the 14x *Phaseolus vulgaris* genome sequence of G19833.**

- ➢ **The positions, alternative alleles, the read coverage and sequence quality at the site of potential SNPs was determined**

- ➢ **SNPs were called both between and within market classes**

# Initial Results of SNP Discovery Using Whole Genome Sequence Analysis

◆ Total number of high quality SNPs among the 19 bean cultivars:

## 1,859,249

◆ The number of SNPs polymorphic in Mapping Populations:

➤ **Stampede X Red Hawk: 241,724**

➤ **BAT93 X Jalo EEP558: 275, 131**

# Initial Results of SNP Discovery Using Whole Genome Sequence Analysis

◆ The following are the number of SNPs within each market class:

➤ **Kidney: 110,163**

**Light Red Kidney: 21,449** ⎤
**Dark Red Kidney: 7,954** ⎦ **776** SNPs polymorphic within both LRK & DRK Market classes

➤ **Pinto: 213,087** ⎤

➤ **Navy: 36,310** ⎬ **687** SNPs polymorphic within all three Market classes

➤ **Black: 49,655** ⎦

# Selection of the Final Set of 6,000 SNPs for the Illumina Infinium GeneChip
## - Distribution Across the 11 Chromosomes -

◆ **The 2nd build of the G19833 whole genome sequence has 10,037 sequence scaffolds.  5,753 scaffolds are 5 kb or greater in length.**

◆ **Identify scaffolds that are heterochromatic**

> **Use the marker distribution in the existing bean genetic maps:**  186 scaffolds (25.4Mb) identified as heterochromatic

◆ **Identify 2 SNPs in the 427 non-heterochromatic scaffolds > 250 kb in length**

◆ **Identify  1 SNP in the remaining 5326 non-heterochromatic scaffolds > 5 kb in length**

# Selection of the Final Set of 6,000 SNPs for the Illumina Infinium GeneChip
## - Polymorphism in Bean Germplasm -

◆ **Must be polymorphic in Stampede x Red Hawk or BAT93 x Jalo EEP558 or both**

◆ **How should SNPs be selected in terms of polymorphism within the various market classes?????**

➢ **All SNPs polymorphic within the LRK + DRK and all SNPs polymorphic within the Black + Navy + Pinto Market classes = 1463 SNPs**

➢ **How should the remainder of the 4500+ SNPs be allocated in terms of**

-  **Mesoamerican vs. Andean market classes ????**

-  **Specific market classes ????**