



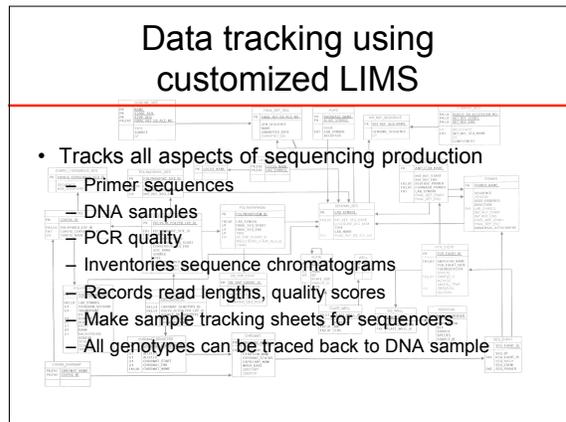
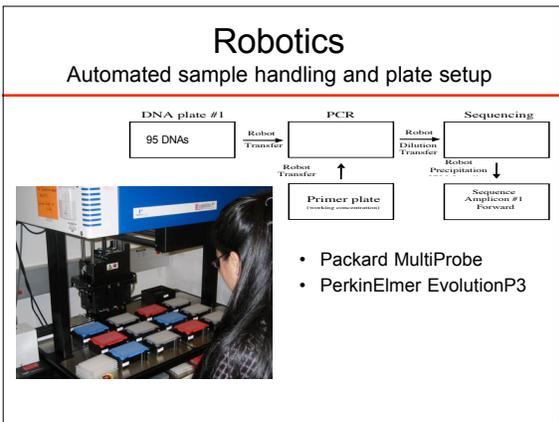
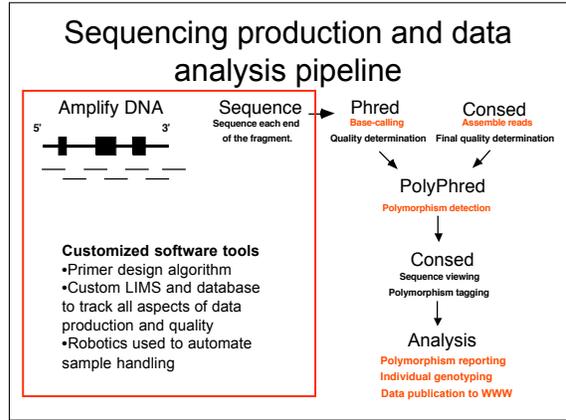
National Institute of Environmental Health Sciences
Environmental Genome Project
NIEHS SNPs

Search Site

NIEHS SNP Tutorial

Department of Genome Sciences
University of Washington

January 30-31, 2006



EGP Bench: Custom LIMS

Organized by gene

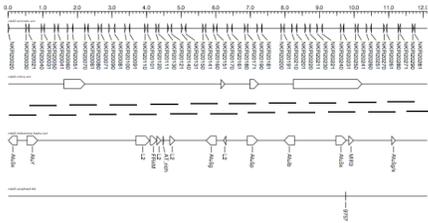
EGP Bench

Generates sample tracking form

Container Name	Plate ID	Description	Application	ContainerType	Owner	Operator	PlateLabeling
Seq-091805.F6	seq_091805.F6		Sequencing/analysis	384-Well	ANALYST_TEST	ANALYST_TEST	Seqts
Well	Sample Name	Comment	Results Group	Instrument	Protocol 1	Instrument	Protocol 2
01	PRAG0000010.4.137623.02		Analysis	Protocol 3	Instrument	Protocol 4	Analysis
02	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
03	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
04	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
05	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
06	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
07	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
08	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
09	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
10	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
11	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
12	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
13	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
14	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
15	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
16	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
17	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
18	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
19	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
20	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
21	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
22	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis
23	PRAG0000010.4.137623.02		Analysis	Protocol 3	Analysis	Protocol 3	Analysis

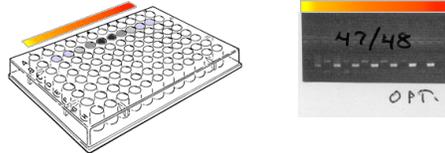
Re-sequencing pipeline

- Gene design- automated primer picking software
 - Exons, 2 kb upstream of first exon, 2 kb downstream of last exon
 - Genes larger than 30 kb have 10% of introns scanned
- Prior to amplification and re-sequencing, problematic GC-rich regions, alu repeats, polynucleotide tracts, and pseudogenes identified



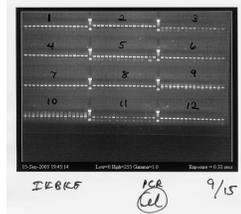
Re-sequencing pipeline

- PCR conditions optimized for each amplicon
- Failed optimization reactions repeated, primers redesigned upon second failure

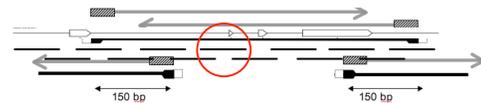


Re-sequencing pipeline

- PCR quality spot checked prior to sequencing
- Failed PCR reactions repeated
- Refractory amplicons subjected to:
 - 10% DMSO
 - Alternate thermocycling parameters
 - Primer redesign

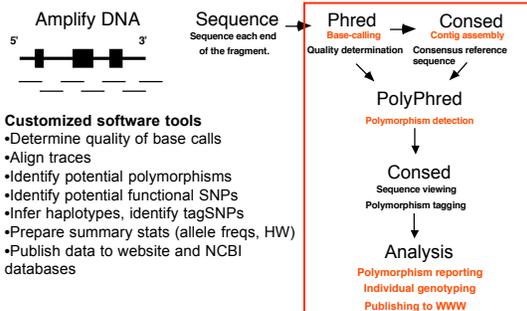


Re-sequencing pipeline

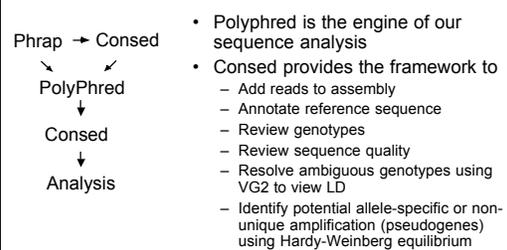


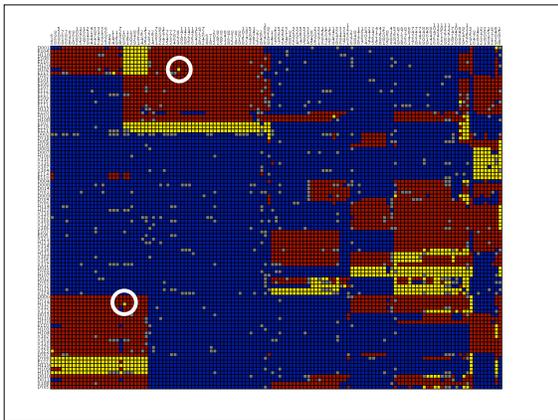
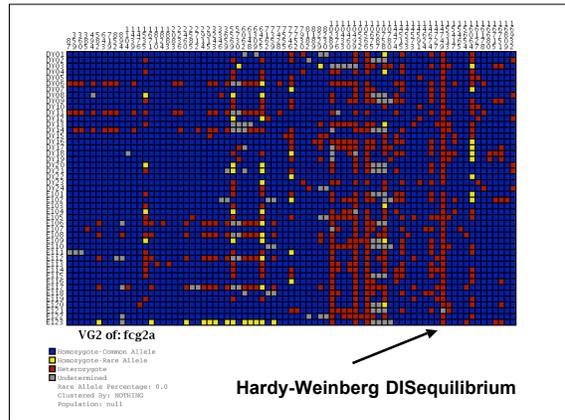
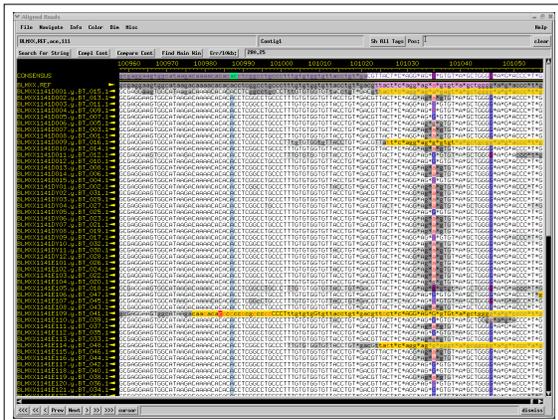
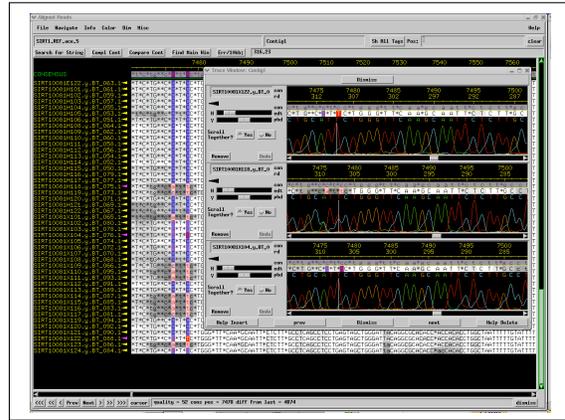
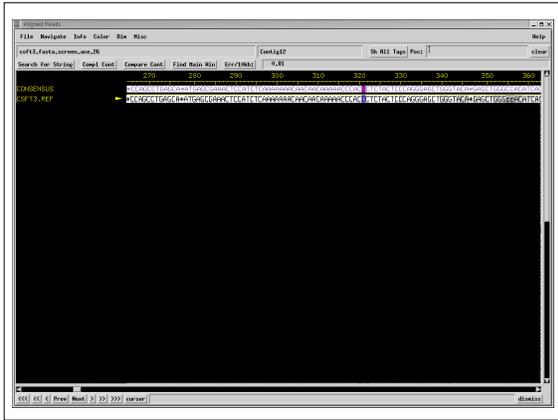
- Standard ABI BDT chemistry
 - Optimized for reaction volume and dilution
- Universal primer sequences standardize sequencing reaction conditions
- ABI 3730 capillary electrophoresis automated sequencers

Data analysis



Sequence Analysis





Extensive Quality Control Protocols and Checkpoints Built into the System

- 1 - Preparation of the reference sequence
 - BLAT analysis of sequence to identify closely homologous regions or pseudogenes
 - Verification of candidate gene mapping and exonic locations
 - Automated entry of baseline sequence and candidate gene information into LIMS
- 2 - PCR primer design
 - BLAST analysis of all primer sequence to ensure specificity
 - Identification of sequence context elements and repetitive sequence which reduce sequencing read lengths and quality
 - Automated entry of all PCR primer sequences and mapping into LIMS
 - Tracking of all PCR primers to candidate gene and ordering information

Extensive Quality Control Protocols and Checkpoints Built into the System

3 – PCR amplification

- Verification of PCR amplification and sizing
- Entry of PCR conditions and PCR results LIMS – linked to specific primers
- Robotic transfer of all DNA samples into pre-made, quality controlled PCR plates

4 - DNA sequencing

- Robotic transfer of all diluted PCR amplicons into pre-made, quality controlled sequencing plates
- Entry of sequencing reaction data into LIMS – linked to specific PCR amplicons and PCR events
- Generation of virtual barcode for each sequencing sample
- Automated generation of sequencing sample sheet (with virtual barcode)
- Daily sequencing reports automatically generated and emailed to laboratory technicians

Extensive Quality Control Protocols and Checkpoints Built into the System

5 - Gene assembly and polymorphism analysis

- Automated entry of sample chromatogram data in LIMS – linked to virtual barcode
- Automated entry of sample chromatogram QC data – Phred quality and read lengths
- Confirmation of orientation and location of sequence data on reference sequence during assembly
- Review of all tagged SNPs by data analyst to confirm quality
- Confirmation of all genotypes using double-stranded data
- Automated entry of polymorphism location and sample genotypes into LIMS

6 - Final data processing

- Confirmation of Hardy-Weinberg equilibrium for all sites (proportion of expected genotypes per site which can reveal problems stemming from allele-specific PCR amplification).

Data publishing

- Text files published to NIEHS SNPs web site and NCBI databases
 - SNP summary data
 - Genotypes
 - Final reference sequence
- Graphical data summaries with GeneSNPs and Visual Genotype images

NIEHS SNPs website

The screenshot shows the NIEHS SNPs website homepage. It features a navigation menu on the left with links like Home, Gene Nomination Form, Gene Targets, Genotyping Resources, Pathway Image Maps, and Personalized. The main content area includes a 'Welcome to the NIEHS SNPs Program' section with an introduction and a 'GeneSNPs Database' section. On the right, there is a 'Latest Updates by Published Genes Table' listing various genes and their update dates.

Data formats published to web facilitate association studies

NIEHS-SPECIFIC LINKS				
Entire Gene	Golden Path (UCSC Genome Browser)	Golden Path (with NIEHS SNPs Tracks)	Pub Med	
Download a zip file of all data for this gene				
Sample Population Description				
Mapping Data	cSNPs cDNA	Color F-ASTA SNP Context	PCR Primers (F-ASTA)	Genbank
Genotyping Data	Visual Genotype Individual Genotypes	SNP Alleles SNP Allele Frequency	SNP Hardy-Weinberg	
Haplotyping Data	PHASE Output Visual Haplotype	Phased Individual Haplotypes	Sorted by Frequency	
Linkage Data	LD Select (Tag SNPs)	African Descent European Descent	Hispanic Descent	Asian Descent
Predictive Analyses	Nonsynonymous cSNP Analysis			

Summary

Amplicons designed to tile across gene region using Tm- matched PCR primers

Amplicons sequenced using standard ABI BDT chemistry

Amplicon sequences assembled into contigs, annotated and reviewed using Consed

Polyphred 5.0 identifies potential SNPs, human reviewed

Custom LIMS tracks all aspects of data production and analysis

Rapid publishing of data files to web and national databases