Borevitz R01GM073822 year1 Progress Report

A. Specific Aims

1. Generation of low and high density haplotype mapping populations.
2. Mapping genetic variation of seasonal response across geographic locations
3. QTL verification, fine mapping, and identification of QTL gene(s)

B. Studies and Results

Several important advances have been made in the first year of research on this proposal. We have nearly completed a broad survey of population genetic variation from local populations in the Midwestern United States, and several locations in Europe including all widely used Arabidopsis accessions available in stock centers. We have developed a SNP/tiling arrays for genotyping 250,000 SNPs, identification of novel Single Feature Polymorphisms (SFPs), and fine genome scanning for Copy Number Variation (CNVs) >150bp. We have performed our first studies of Seasonal Variation under simulated laboratory conditions (LI *et al.* 2006).

This work is being spearheaded by an excellent quantitative genetics postdoc Dr Yan Li, who is supervising technicians and undergrads in the Borevitz lab. Collaboration with Nordborg lab at USC has been intimate in developing the SNP/tiling array (GILAD and BOREVITZ 2006; SHIU and BOREVITZ 2006; ZHANG *et al.* 2007) and in analyzing local population genetic structure. Work with the Marjoram lab has calibrated methods for scanning the genome for diversity and recombination using SFP data (JIANG *et al.* 2006). Work with the Nordborg and Marjoram labs will proceed via methods for whole genome association mapping recently established (ZHAO *et al.* 2007). Work with the Zoellner lab has developed methods for detection and frequency estimation of SFPs and CNVs jointly across multiple samples using Hidden Markov Models.

Our field work in year 1 has been very fruitful. We have collected 1000s of lines and invited the community to donate their own collections. The rapid genetic fingerprinting has come in under budget, which allows us to increase sample size including profiling the entire stock center collection (re-numbered to match the current public genotypes). Field sites contain genetic resources: natural mutation accumulation lines, natural recombinant inbred lines, natural near isogenic lines, natural reciprocal transplant experiments and natural longitudinal fitness studies. Some populations are monomorphic at 149 common SNPs but contain phenotypic variation likely due to new mutations, while other sites contain multiple strains adapting by recombining preexisting diversity. Abiotic environmental signals will guide our experiments in the lab under controlled conditions that mimic local seasonal variation.

The new *Arabidopsis* SNP/Tiling Array (AtSNPtile1)

Whole genome tiling arrays are versatile tools to read out various forms of genomic data (BOREVITZ and ECKER 2004). SNP arrays are also widely used, and the 100k and 500k SNP arrays are now popular in humans. These arrays contain multiple probes specifically designed to assay known SNPs. The genotype call is made by contrasting intensities from probes of different alleles. Several robust statistical methods

have just been released, including RLMM (RABBEE and SPEED 2006), BRLMM, and CRLMM. The last Arabidopsis SNP array, which we used to characterize linkage disequilibrium, contained a meager 412 SNPs (NORDBORG *et al.* 2002). Recently, Detlef Weigel & coworkers (Max Planck Institute, Tübingen, Germany), in collaboration with Perlegen Sciences, Inc., have used whole genome resequencing arrays to identify more than 650,000 high confidence SNPs in 20 diverse *Arabidopsis thaliana* accessions (Richard Clark et al., manuscript under review). Using this unparalleled resource, the PI and M. Nordborg (USC) have developed a 2nd generation SNP/tiling array for *Arabidopsis thaliana* **(AtSNPtile1).** This array contains probes for each allele and each strand for **250,000 common SNPs and 1.7 Million unique 25mer tiling probes** covering the nonrepetitive part of the genome at ~35 bp resolution. More than 100,000 SNPs map to transcribed regions of the genome. Importantly, 22,923 coding genes contain 1 or more SNPs on AtSNPtile1, while 18,767 contain 2 or more SNPs, thereby allowing comprehensive investigation of allele specific expression. AtSNPtile1 also contains probes to centrally cover all 130,000 CCGG sites for methylation analysis. We have already received 1400 AtSNPtile1 arrays to develop 3 core Haplotype Mapping populations under Borevitz NIH R01GM073822 and Nordborg NSF 2010 #0519961. AtSNPtile1 is now publicly available from Affymetrix.

This year we released our first publication "Genetics of Local Adaptation in the Laboratory" where we mapped QTL in recombinant inbred lines for flowering time under spring conditions from geographic location spanning the native range of the species. Major and minor QTL were detected that interact with each other and the environment, demonstrating context dependence phenotypic effects expected for loci with adaptive importance. SolarCalc software was developed to run chamber settings at arbitrary GPS coordinates world wide. Light intensity, light spectrum, humidity and temperature setting mimic the 30 year average for any site at 1 minute intervals. This control will allow fine ecological parameters affecting plant growth to be dissected in laboratory settings across diverse HapMap genotypes.

C. Significance

This proposal is on track to establish the methods and genetic resources for whole genome linkage disequilibrium mapping in *A. thaliana.* Taking advantage of the genetic toolkit of this model organism will also us and the greater community to verify and clone QTL detected by association mapping. In this way alternative haplotype based mapping methods that account for population structure can be compared and linkage validated empirically. Such direct tests will be a boon to studies in humans where linkage strength is often the only signal with which to separate false positives from true. In addition, general conclusions about the genetic architecture of adaptive traits in widely distributed and exponentially expanding populations can help set reasonable models for experimental designs in other species.

These tools will draw in the broader molecular biology and emerging ecological *Arabidopsis* community to define evolutionary mechanisms underlying physiological and developmental processes of adaptation. Broader ecological issues relating to invasive species are emerging as strains are distributed by humans colonizing new disturbed sites under similar agricultural programs. Furthermore climate change is allowing the

northward flow of late fall germinating lines. *A thaliana* may prove to be a begin genetic marker of this progression.

D. Plans

1) We will select a set of 384 maximal diversity accessions from across the world to genotype at 250,000 SNPs to provide maximums power for whole genome LD mapping.
2) This panel will be grown in simulated spring and fall seasons from north and south latitude ranges to map QTL for local environmental responses.
3) Select QTL loci will be confirmed in F2 crosses, followed by fine mapping and ultimately QTL gene identification.

Human subject

No human subjects are involved; it is a plant genetics project.

Vertebrate animals

No Vertebrate animals are involved; it is a plant genetics project.

E. Publication list

BOREVITZ, J. O., and J. R. ECKER, 2004 Plant genomics: the third wave. Annu Rev Genomics Hum Genet **5:** 443-477.

GILAD, Y., and J. BOREVITZ, 2006 Using DNA microarrays to study natural variation. Curr Opin Genet Dev **16:** 553-558.

JIANG, R., P. MARJORAM, J. O. BOREVITZ and S. TAVARE, 2006 Inferring population parameters from single-feature polymorphism data. Genetics **173:** 2257-2267.

LI, Y., P. ROYCEWICZ, E. SMITH and J. O. BOREVITZ, 2006 Genetics of Local Adaptation in the Laboratory: Flowering Time Quantitative Trait Loci under Geographic and Seasonal Conditions in Arabidopsis. PLoS ONE **1:** e105.

NORDBORG, M., J. O. BOREVITZ, J. BERGELSON, C. C. BERRY, J. CHORY *et al.*, 2002 The extent of linkage disequilibrium in Arabidopsis thaliana. Nat Genet **30:** 190-193.

RABBEE, N., and T. P. SPEED, 2006 A genotype calling algorithm for affymetrix SNP arrays. Bioinformatics **22:** 7-12.

SHIU, S. H., and J. O. BOREVITZ, 2006 The next generation of microarray research: applications in evolutionary and ecological genomics. Heredity.

ZHANG, X., E. J. RICHARDS and J. BOREVITZ, 2007 Genetic and epigenetic dissection of cis regulatory variation. Current Opinion in Plant Biology **10**.

ZHAO, K., M. J. ARANZANA, S. KIM, C. LISTER, C. SHINDO *et al.*, 2007 An Arabidopsis example of association mapping in structured samples. PLoS Genet **3**(1): e4.

F. Project-generated Resources

Web resources

We currently posted genetic data at low resolution for >1700 Accessions at 149 framework SNPs
We have pictures at flowering for those same accessions
We have flowering time Data for a single Long Day experiment

We have preliminary analysis scripts for population structure
We have the first hybridizations for AtSNPtile1
We have supplementary data for Li et al, 2006
We have a season calculator

All is publicly available here and free.
http://naturalvariation.org/hapmap