1. Introduction thoughts (from lab meeting)
   1. Intraspecific variation within Botrytis cinerea alters both pathogen virulence and plant susceptibility
   2. Mechanism: B. cinerea genetic control of gene expression
      1. And B. cinerea genetic control of host gene expression
   3. Driving questions
      1. How does the genome of B. cinerea relate to the transcriptome of A. thaliana?
      2. How does the genome of B. cinerea relate to the transcriptome of B. cinerea?
      3. What genetic variation in B. cinerea determines gene expression in infected A. thaliana?
      4. What genes in A. thaliana are sensitive to natural variation in B. cinerea?
2. Analysis ideas
   1. Alternate approach for eQTL: R package matrixeqtl?
      1. install.packages("MatrixEQTL")
   2. Plot of SNP density along genome
   3. Plot of transcript density along genome
   4. Test hotspots: are transcripts more tightly coexpressed within hotspots than between?
      1. For BOA: statistically AND biologically enriched
   5. Use top eQTL hotspots as factors in linear models of gene expression
      1. For all genes with significant expression change due to focal eQTL, evidence of a shared regulatory gene network under eQTL (TF/ master regulator)?
      2. How many expression profiles are correlated with each hotspot?
      3. What is the magnitude of phenotypic effect of the eQTL hotspots?
   6. Idea (maybe, I’m tired): by following gene-gene connections (hotspot SNP/gene 🡪 linked transcripts 🡪 secondary transcripts linked to these genes can I sort of build a rough gene regulatory network? Or is that \*way too wild and sloppy\*?
      1. Could select a couple of favorite hotspots and try this approach
      2. Could filter for hotspot overlap with Wei’s networks
   7. From Celine Eudicot paper: once eQTL genes are identified, compare them to the secreted proteins list. If secreted, these eQTL (ESP if eQTL on Arabidopsis) could be interpreted on having a direct effect!
      1. Celine: “proteins (Table S3) identified in the Botrytis secretome (49.3% of the secreted proteins successfully identified in the B05.10 genome), suggesting the importance of secreted proteins in the successful virulence across plant species (Gonzalez-Fernandez et al. 2015).”
      2. Celine: “SNPs were annotated based on their location in ASM83294v1 assembly while gene annotation was extracted from the fungal genomic resources portal (fungidb.org). The Botrytis secretome data including 1220 proteins was extracted from the Fungal Secretome Database (fsd.snu.ac.kr).”
   8. Try Runcie analysis for major ~~PCA factors driving expression variation
3. Interpretation ideas
   1. Look for overlap between my hotspots and Zhang 2017/ 2018 genes
   2. Interpret Bc hotspot peaks with no At hits:
      1. Bc expression changes, but does not cause changes on this host (col0)
   3. Check for gene IDs in BOA cluster: last 2, different pathway? Next 2, downstream modifications?
   4. Collapsing 11 organics into 1 “average” individual
      1. Expect: mostly don’t matter
         1. Either have the major allele & phenotype effects are swamped by ~40 other individuals
         2. Or have minor allele alone and will be dropped at 20% MAF
      2. Or: maybe sometimes have minor allele with other and could find some peaks due to overrepresented SNPs
   5. Thresholding from permutation
      1. No single threshold works: when permutation p-values are small, observed data also has small p
      2. 100% threshold is too conservative for most genes
4. Meta-analysis:
   1. GWAS (Bc virulence on Col0) vs. eQTL hotspots
   2. transcriptome (Bc expression on Col0 // Col0 expression) vs. eQTL hotspots
   3. filter genes (transcripts) for those with GWAS hits on a plant host. THEN look for eQTL hotspots (with a phenotypic connection)
      1. on At Col0
      2. on tomato
   4. GWAS vs. transcriptome
5. Chen 2015 Plant pathogen gene networks
   1. This, but also Ref 7,8: host and pathogen coexpression analysis
      1. 8: candida and mouse, Tierney
      2. Marguerat S, Bähler J. RNA-seq: from technology to biology. Cellular and molecular life sciences. 2010; 67(4):569–79. doi: 10.1007/s00018-009-0180-6 PMID: 19859660
      3. Tierney L, Linde J, Müller S, Brunke S, Molina JC, Hube B, et al. An interspecies regulatory network inferred from simultaneous RNA-seq of Candida albicans invading innate immune cells. Frontiers in microbiology. 2012;3. doi: 10.3389/fmicb.2012.00085 PMID: 22416242
   2. Poplar and Marssonina brunnea
   3. One possible mechanism of DIRECT Botrytis eQTL on host: if eQTL is a secreted effector
6. Morel 2018 Grape diversity paper
   1. Genetic diversity in ~ all agricultural samplings except CA? (ours, Ma & Michailides 2015)
   2. Weak association pop str with geography
   3. Attempted crosses and assessed morphotypes!
      1. Methods to determine mating type by PCR
   4. Terminology: “aggressiveness” for “virulence”
   5. Independent test strategy to build an interspecies network topology
      1. Crazy to do this for the eQTL idea?
7. Samad-Zamini 2017 plant pathogen eQTL paper
   1. eQTL may be independent of phenotypic QTL
   2. eQTL may be time-dependent in the plant-pathogen interaction
8. Perez-Enciso 2007 genetical genomics eQTL cofactors
   1. Use all transcripts as potential regressors in eQTL analysis of each transcript
   2. it may not be possible to disentangle between a direct effect of the marker on the gene or an indirect (spurious) effect caused by an intermediate gene in the same or co-regulated metabolic route.
      1. This is fine if we only care about eQTL hotspots controlling many loci
9. Guo 2017 symbiosis paper
   1. eQTL hotspots may be pleiotropic—one gene modulating expression of many genes
      1. interspecific pleiotropy!
      2. From paper “Host Expression Modulator” for parasite loci
   2. Could add a co-expression analysis across host and pathogen?
      1. Gene co-expression with isolate genotype as the independent variable
      2. Gene co-expression with isolate genotype \*at specific loci\* as the independent variable
   3. Evaluate effect size of eQTL on each gene?
      1. How can I calculate the predicted effect of the single top SNP hit locus on gene of interest? Include SNP state within a block as the factor?
   4. Overrepresentation analysis of eQTL hotspots
   5. Look at transcripts modulated by a hotspot
      1. Overrepresentation analysis in functions of transcripts affected by a hotspot!
      2. // function of genes in hotspot
   6. Overrepresentation of which genes have significant cis effects?
   7. Network analysis of polymorphic genes and affected transcripts?
   8. Can I calculate LOD scores? Goal: find nearby loci with LOD scores in opposite directions
   9. Recombination map?
      1. Can tell multiple vs. single causal locus per eQTL based on clustering of recombinant vs. nonrecombinant individuals within region of eQTL
   10. Interpretation of local eQTL: “genes for which expression/ mRNA abundance is strongly associated with genotype near their genomic position”
   11. Prominent vertical bands: “genomic loci that influence the expression of genes located throughout the genome” = trans-eQTL hotspots
10. Pei 2018 Kiwifruit Botrytis
    1. Geographical origin did not predict pathogenicity
    2. “relatively uniform species diversity”
    3. Phenotype & genotype variation/ differentiated within population
    4. Transposon study—boty, flipper – structure into subgroups : both/ b/ f/ none
       1. Transposa (both) may be more virulent than vacuma (neither)
       2. And temperature-sensitive
    5. Sexual compatibility: MAT1-1 and MAT1-2 ideomorphs at one gene—if both, then self-fertile.
       1. If 1:1, assume random mating and frequency-dependent selection
    6. Frequency of different morphologies: mostly aerial hyphae, produce conidia, produce sclerotia
    7. Disease more severe on young than mature leaves
11. Hernandez 2012
    1. Method for cis eQTL detection:
       1. Per SNP, take all transcripts within 500kb, incorporate them into linear model. Estimate “association between the allelic does of each SNP as a predictor of proximal gene expression levels”
12. Atwell 2018 Bc genome (T4, mostly)
    1. Genome-wide map of LD?? Try controlling for this in some analysis // compare to hotspots
       1. LD decay
       2. LD value across genome
    2. Similarly, LOF mutations and selective sweeps genome-wide?
    3. Some way to include isolate grouping in results?
       1. Refer to STRUCTURE outputs! K=4 major groups. Isolates from group 1 or admixed group 2-4
       2. Phylogeny on neighbor-joining tree
       3. How to deal with organics.
    4. GWA with additional polymorphism information
       1. mitochondrial SNPs
       2. indel polymorphisms
    5. SNPs with major LOF polymorphism – check SNP lists for these
    6. Concern: how do 11 “identical” organic isolates weight the GWA? Will they tend to disproportionately find associations?
       1. I assume a relatedness matrix (k-matrix) would address this somewhat—READ UP
13. Nicolae 2010 GWAS eQTL
    1. eQTL hotspots = master regulators
    2. eQTL hotspots (SNPs associated with many transcripts) are enriched among trait-associated SNPs… aka trait-associated SNPs are likely eQTLs
       1. may only be true for cis-eQTL, not so much trans-eQTL
       2. more true the more stringent the threshold is for calling a SNP “trait-associated”
       3. my analysis: a) identify eQTL hotspots (in progress)
       4. b) check whether these SNPs are \*enriched\* / overlap with SNPs associated with B. cinerea virulence on A. thaliana or eudicots
    3. simulation method to test robustness of eQTL calculation
    4. how to define “cis” within a chromosome?
       1. In humans, 4 Mb
    5. “true associations may be more easily detected in regions of the genome with high LD because these regions are more likely to have good coverage on high throughput platforms for GWAS”