Functional gene categories differentiate maize leaf drought-related microbial epiphytic communities

Barbara A Methe[1,2], David Hiltbrand[3], Jeffrey Roach[4], Wenwei Xu[5], Stuart G Gordon[6], Brad W Goodner[7], Ann E Stapleton[3],

1 J Craig Venter Institute, Rockville, Medical Center Drive, 20850 Rockville, MD, USA **2** Department of Medicine, University of Pittsburgh, 3459 Fifth Ave, 15213 Pittsburgh, PA, USA **3** Department of Biology and Marine Biology, University of North Carolina Wilmington, Wilmington, NC, USA **4** 4Research Computing, University of North Carolina Chapel Hill, 201 South Columbia Street, 27514 Chapel Hill, NC, USA **5** 5Agricultural and Extension Center, Texas A and M AgriLife Research, 1102 East FM 1294, 79403 Lubbock, TX, USA **6** Biology Department, Presbyterian College, 503 South Broad Street, 29325 Clinton, SC, USA **7** Department, Hiram College, 11715 Garfield Road, 44234 Hiram, OH, USA

*stapletona@uncw.edu

Abstract

The phyllosphere epiphytic microbiome is composed of microorganisms that colonize the external aerial portions of plants. Relationships of plant responses to specific microorganisms–both pathogenic and beneficial–have been examined, but the phyllosphere microbiome functional and metabolic profile responses are not well described. Changing crop growth conditions, such as increased drought, can have profound impacts on crop productivity. Also, epiphytic microbial communities provide a new target for crop yield optimization. We compared *Zea mays* leaf microbiomes collected under drought and well-watered conditions by examining functional gene annotation patterns across three physically disparate locations each with and without drought treatment, through the application of short read metagenomic sequencing. Drought samples exhibited different functional sequence compositions at each of the three field sites. Maize phyllosphere functional profiles revealed a wide variety of metabolic and regulatory processes that differed in drought and normal water conditions and provide key baseline information for future selective breeding.

Introduction

Plants form a wide variety of intimate associations with a diversity of microorganisms in the phyllosphere, the above-ground plant surface [1, 2]. Microorganisms can exist as endophytes within the plant, as epiphytes on plant surfaces (which together compose the phyllosphere) and in the soil surrounding and in the roots [3–5]. The ubiquity and intricacy of these plant-microbe associations support the model of the plant as a "meta-organism" or "holobiont" consisting of the host and its microbiome (the collection of microorganisms and their gene content) which maintain a relationship over the lifetime of the plant [3, 6, 7]. The plant-associated microbiome, the phytobiome, is a complex and dynamic system existing as both an agonist and antagonist of plant fitness and adaptability [8, 9]. Therefore, elucidating the nature and extent of these interactions

offers significant opportunities for improving plant health, for example, through alterations in nutrient cycling, neutralizing toxic compounds, discouraging pathogens, and promoting resistance to abiotic stresses that have the potential for generating significant impact on plant productivity [10–14]. Optimization of selective breeding for epiphytes presents new challenges in ensuring that microbe colonization occurs as needed, while presenting new potential effective indirect genetic selection [14–16] for crop improvement. Ultimately, engineering microbial and plant genotypes for optimal function and resilience will also require causal, mechanistic analyses of gene and pathway level processes; one first step in such mechanistic analysis for the microbial components of the phyllosphere is construction of controlled synthetic communities of microbes or assembly of specific sets of microbial functional genes [17].

In contrast to the rhizosphere, the region of soil that is directly influenced by root secretions, the phyllosphere is both a relatively understudied and transitory microbial environment [2, 18]. Microbial epiphytes of the phyllosphere experience an environment subject to different influences than those found in the rhizosphere and from host endophytes. Those in the phyllosphere experience atmospheric influences including direct sunlight exposure during diurnal cycles, and barriers such as waxy cuticle resulting in an oligotrophic environment [19, 20]. More labile associations between epiphytic microbes and host leaves do present an opportunity for interventions. For example, inoculation of beneficials or application of probiotics [21] could be done rapidly, during crop growth, since above-ground leaves and stems are easy to access. Longer term interventions such as selection of host genotypes that support specific desired microbial functions on external leaf surfaces at key points during growth or in response to biotic or abiotic stress could also be attempted [22, 23].

Corn, *Zea mays* L., is a widely grown and economically important annual crop. Drought is an abiotic stress that can negatively affect plant productivity [24]. Hence, understanding the role that the phyllosphere may play in association with maize undergoing abiotic stress is a priority. Epiphytic microbes are a unique target for drought tolerance. Targeting such microbes has potential advantages in the speed of alterations relative to plant breeding. It also provides the potential for temporal targeting through inoculation only during the adverse conditions [14]. Supporting this potential, seed microbial inoculation for crop drought tolerance is already in commercial use (for example, https://www.indigoag.com/).

Only between 0.1% and 10% of the microbial diversity is culturable in vitro [25, 26]. This has led to the use of culture independent methods for study of microbial community structure and function. For approximately the past two decades, microbial and fungal diversity has been described via the sequencing of amplicons representing biomarkers, such as the 16S rRNA gene (bacteria) and internally transcribed spacer (ITS) regions (fungi) [27, 28].More recently, techniques in microbial community research have shifted to investigate community structure and function at a systems-wide level. One such systems method, metagenomics, involves sequencing and analyzing genes derived from whole communities as opposed to individual genomes. Examining microbiomes at this level has shown that microbes ultimately function within communities rather than as individual species [9]. The traditional use of taxa to investigate microbiomes does not fully account for metabolic interactions between species. Typically functional genes exhibit different patterns than taxa, and functional genes are often better predictors of niche [29–31]. In addition, functional gene content can be more heritable (i.e., more driven by host genetic interactions) [32]. Functional gene analyses also provide key information needed for community-level metabolic engineering [14, 22].

To address our questions about functional differences between microbial communities, we selected a factorial design with use of multiple field sites to increase generality. We know that plant breeding requires consideration of environmental

contributions. By prioritizing multiple field sites in our initial investigations, our results provide critical information for future experimental designs for breeding and extension of the experiments found here.

Seed Stocks

Zea mays L. inbred B73 seed was supplied by the Maize Stock Center, http://maizecoop.cropsci.uiuc.edu/, and seed was increased at the Central Crops Agricultural Research Station, Clayton, NC using standard maize nursery procedures. Genotype of the seed lots used for these experiments was verified by SSR genotyping using eleven markers and comparison of fragment sizes to the sizes listed in the MaizeGDB database [33], http://www.maizegdb.org/ssr.php.

Experimental Design and Field Sampling

Research field sites were generously provided by collaborators with ongoing scientific and extension projects; no additional permits or permissions were required. For this experiment we used a hierarchical design, with the treatment plots nested in each field site. There were three randomly arranged plots within each treatment level at each field site, surrounded by additional plant plots. Replicated field plots were planted in Albany, CA at the USDA University of California-Berkeley field site (abbreviated as CA), 37 degrees 53 min 12.8 sec N 122 degrees 17 min 59.8 sec W, on June 6, 2012. The field site had uniform soil and subsurface irrigation and fertilization supplied according to normal agronomic practice for this growing area. The southern section had normal irrigation throughout the season. However, the northern section had normal irrigation until vegetative growth stage V5 when all watering was stopped for two weeks; after sampling of leaves irrigation was resumed to allow plant growth to maturity. Seeds were planted at two sites in Texas, Dumas Etter field (abbreviated as DE) 35.998744 degrees N 101.988583 degrees W on May 8, 2013, and Halfway, TX field (abbreviated as HF), 34.184136 degrees N 101.943636 degrees W on April 26, 2012. The sites had center-pivot irrigation and standard maize field management. Drought treatment blocks were watered at 75% of the normal rate at DE and at 50% of the normal rate at the HF field site. The DE field site had one replicate plot that experienced additional rain late in the season (after phyllosphere sample collection). The HF field site had no unmanaged precipitation between July 9 and harvest. Late-season (post-phyllosphere sampling) field trait measurement methods and data files for each field site are provided in Supplemental Plant Traits files 1-7 in synapse repository syn15575565.

Field Trait Measurement

At the Texas field sites (DE and HF) plant and ear heights were measured once per plot when growth was complete after tasseling. Ears were harvested and shipped to UNCW for measurement. For each ear, cob diameter at the base was measured with digital calipers, and twenty seeds were removed from the middle of each cob, placed in envelopes, and weighed. For the CA site, individual plant heights were measured and cobs were collected at the end of the season, October 1–3, 2012. Seed development was not complete, so only cob traits were measured. Cob diameter at base was measured with digital calipers; cob length was measured with a ruler. Plant data for each location and trait are included in syn15575565 as plant trait files 1 to 6, with metadata about the column headers in file 7.

Leaf Sampling and DNA Extraction

Samples were collected from DE on June 26, 2012 and from HF on June 27, 2012, at developmental stage V8. The CA phyllosphere samples were taken August 7 and 8, 2012, at developmental stage V8. Six fully expanded leaves from the top quarter of the plants in each plot were placed into sterile bags (Whirl-Pak, Nasco, Fort Atkinson, WI) prefilled with 300 mL sterile water and 3 microliters Silwet L-77 (EMCO, North Chicago, IL). Bags were moved to nearby shelters, sonicated for one minute to loosen epiphytic microbes, and the 300 mL of wash solution was filtered through sterile Pall microfunnel 0.2 micron filter cups (VWR, Radner, PA) to collect microbial cells on the filter surface. The filter was removed from the cup with sterile tweezers and dropped into small sterile Whirl-Pak bags then stored frozen until DNA extraction. DNA was extracted from each filter with a PowerSoil Mega kit (MoBio, Carlsbad, CA). Samples were concentrated with filter-sterilized sodium chloride and absolute ethanol according to the manufacturer's instructions and shipped frozen to JCVI for sequencing. Supplemental methods video links are available in the supplemental files repository syn15575565, to provide additional details on the protocol used for leaf washes and filtering.

Library Construction and Sequencing

All library construction and sequencing were completed using Illumina reagents and protocols. Samples PHYLLO09 and PLYLLO10 were sequenced with Illumina HiSeq and all other samples were sequencing using the MiSeq platform. Two nucleic acid negative control filters were also processed through DNA extraction and library construction and sequencing to test for the presence of any significant contamination of experimental samples by exogenous DNA. One sample from HF drought was lost during processing. The raw data and processed reads are accessible from the NCBI Short Read Archive under Bioproject PRJNA297239.

Because of the nature of the sampling collection and nucleic acid procedures, plant host genomic DNA was inevitably included in the nucleic acid samples used for library construction. Therefore, a screening process was implemented to remove both sequencing artifacts and reads most likely to be of maize origin. Adaptor sequences were removed from the SRA sequencing reads using Trim Galore version 0.4.3 <https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/>. The reads were subsequently filtered to remove maize sequences by alignment using bowtie2 version 2.2.9 [34] to v4 of the B73 Zea mays reference, Zm-B73-REFERENCE-GRAMENE-4.0 <ftp://ftp.ensemblgenomes.org/pub/plants/release-37/fasta/zea_mays/dna/Zea_mays.AGPv4.dna.toplevel.fa.gz> [35]. Quality control at each processing step: initial reads, after adapter trimming, and after host filtering, was verified by FastQC v0.11.7. Read pairs that could be joined were joined with vsearch, v1.10.2_linux_x86_64, <https://github.com/torognes/vsearch> [36] and all resulting single-end reads: those that joined and those that did not, were retained for further analysis. UniProt50 protein annotation was preformed by HUMAnN2 v0.9.1,<https://github.com/leylabmpi/humann2>, [37] resulting in estimates of gene family count, path abundance, and path coverage together with estimates of taxonomic profile at the species level generated by MetaPhlAn2 [38]. Gene family HUMAnN2 output was explicitly normalized to counts per megabase to adjust for different input library sizes. Full details of parameters, software packages, and scripts used to manage analyses are available in synapse repository syn12933189.

Count Data Analysis

Analysis of the number of reads for each UniProt annotation in each sample was performed with ENNB [39]. The parameters and full R scripts for analyzing the data (along with an R notebook explaining the process) are available in synapse repository syn12933189. ENNB is a two-stage process with an elastic net for feature selection then negative binomial fit to identify significant annotations, though it is only possible to fit one factor (nested or full factorials for multiple experimental factors are not possible to fit using this two-stage multivariate method). The package was downloaded from the An web page (http://cals.arizona.edu/ anling/software.htm) and scripts written to run both method 1, the trimmed mean (TMM) from the EdgeR package, and method 2, DE-Seq-type count overdispersion. Statistical analysis of annotations different in drought and well-watered conditions were carried out for each field site. The missing HF and CA samples were imputed using the R package MI for the ENNB analysis. After analysis, the annotation data sets were cleaned to remove any rows with annotation IDs that were present in the soil or mock-collected sequenced samples. All input files, R code, an R notebook explaining the analysis, and output files are available at the syn12933189 repository.

Visualization of Significant Annotations

Uniprot lists were converted to Gene Ontology lists (not a 1 to 1 mapping) using the conversion web tool at EBI, with lists available in the supplemental data in syn12976174, then the lists of GO Process and GO Function annotations that were significantly different upon output from ENNB were visualized using REVIGO [40], http://revigo.irb.hr/, with the Simrel and medium list defaults selected. The REVIGO cytoscape-format xgmml network files were color-coded and the network layout redrawn using Cytoscape v3.2.1 [41] . Venn diagrams for comparison of lists were created with http://www.webgestalt.org/GOView/ [42].

Simulation Construction for Analysis Validation

In order to measure the precision and accuracy of our analysis pipeline, we constructed simulated files of sequences and processed these through our analysis pipeline to generate simulated counts. Then, we analyzed the simulated counts with ENNB and functions to tabulate true and false positives. We modified and updated FunctionSim (https://cals.arizona.edu/ anling/software/FunctionSIM.htm) to generate sequences with signal and noise that were made independently of our real data. The full set of scripts and parameters is available in synapse repositories syn15575560 and syn15575551. We tested multiple ENNB thresholds for declaring significant annotations to select suitable cutoff and analysis options with the lowest possible false positive rate.

Statistical Analysis of Plant Traits

Plant traits (seed weight, plant height, and cob diameter) were analyzed with linear regression models using JMP11 Pro (SAS, Cary, NC) with an alpha of 0.05. Models were fit with water treatment (as a nominal factor) for each trait. For HF and DE cob diameter traits, plot numbers were used to identify the group of plants within the larger field and those plot IDs were included in the model to account for the blocks. The number of replicates for each comparison is provided in the box plot figure legend.

Availability of Data and Materials

Metagenomic sequences are available in the SRA repository, identifier BIOPROJECT PRJNA297239. All data analysis scripts, simulations, intermediate files and metadata files are available in seven `synapse.org` repositories, listed by folder title and identifier: 1) Supplemental figures and files syn15575565 (with doi:10.7303/syn15575565),

2) GEN-SAMPLES-2 syn15575562 (with doi:10.7303/syn15575562.1), 3) SimSamples syn15575560 (with doi:10.7303/syn15575560.1), 4) DNApolymAIII syn15575551 (with doi:10.7303/syn15575551.1),

5) DELIVER-FILTER syn15575484 (with doi:10.7303/syn15575484), 6) GOanalyses syn12976174 (with doi:10.7303/syn12976174),

7) Drought_metagen syn12933189 (with doi:10.7303/syn12933189). A preliminary version of this work is available in bioRxiv under bioRxiv 104331 doi `https://doi.org/10.1101/104331`.

Results

To examine the microbial metabolic and regulatory functions important for leaf epiphytic community differences between drought and well-watered field plots, we developed a nested experimental design and a per-field-site analysis using factorial multivariate approaches suitable for zero-inflated annotation read count data. We prioritized comparisons within multiple geographically diverse field sites. Genotype–environment interaction is a key logistic and experimental constraint for future host plant breeding for improved varieties that would support optimal microbial communities.

We saw little correlation between depth of microbial sequence and annotation quality (Table 1). Both of the soil samples and one mock sample had no sequence signal (Table 1). The second mock sample contained some sequences that were not classified as contaminants. All annotation rows present in the mock sample were removed from all sample rows before statistical analysis.

Annotations Differing Between Drought and Watered Treatments

To robustly determine the ENNB parameters with the fewest false positives we created simulations using an independent sequence database. Then, we processed the simulations through our sequence read and statistical analysis code and measured the number of true and false detections. For count analysis, use of the trimmed mean adjustment (Tmm1) and a threshold of 0.001 for negative binomial fitting gave fewer false positives (supplemental repository folder syn12933189) and we report results using those thresholds. Our analyses may be re-run using the scripts and setting information provided in the synapse repository supplemental files if different thresholds are desired. Drought and watered plots at each site site had significant differences in read counts for regulatory and metabolic functions. The ENNB analysis with normalization by TMM generated a list of significant GO Process and GO Function annotations in watered as compared to drought-treated phyllosphere samples for each field site, with groups of related GO terms from REVIGO analysis indicated by edges between GO node terms. Larger nodes indicate the frequency of the annotation in the GO database, so smaller nodes with no edges such as bacteriocin immunity (Fig 1a) are the most unique. The significant GO Process terms identified as semantically distinct in the drought treatment for the Albany, CA field (abbreviated as CA) site (Fig 1a) include biochemical pathways involved in basic cellular responses, such as transcription and DNA replication, and specific metabolic remodeling pathways, such as isoleucine biosynthesis. Pathways we observed that are likely to be important for microbial community interactions include

Table 1. Sample Characteristics

Sample ID	Field Site[1]	Treatment Type	Sequence Amount[2]	Sequence Comment
PHYLLO9	HF	watered	deep	
PHYLLO10	DE	watered	deep	
PHYLLO11	CA	watered	small	all contaminant
PHYLLO12	CA	drought	large	low proportion of signal
PHYLLO13	CA	watered	large	
PHYLLO14	CA	drought	moderate	
PHYLLO15	DE	watered	moderate	
PHYLLO16	DE	drought	small	
PHYLLO17	CA soil	watered	small	soil sample below watered plot plants, all contaminant
PHYLLO18	CA soil	drought	small	soil sample below drought plot plants, all contaminant
PHYLLO19	mockDE	none	small	
PHYLLO20	mockCA	none	small	low proportion of signal
PHYLLO21	CA	watered	large	
PHYLLO22	CA	drought	large	
PHYLLO23	DE	watered	large	
PHYLLO24	DE	drought	small	
PHYLLO25	DE	drought	moderate	
PHYLLO26	HF	watered	deep	
PHYLLO27	HF	watered	deep	
PHYLLO28	HF	drought	deep	
PHYLLO29	HF	drought	deep	

[1] Full field information for these two-letter abbreviations is available in the Methods section.
[2] Small indicated that the sample contained less than 233k reads, moderate indicates 233-500k reads, large indicates 500k-1.6m reads, deep indicates greater than 1.7m reads.

bacteriocin immunity and amino acid transport [43]. Functional annotations (Fig 1b) for the CA field site are similar to process annotations, with the addition of a cluster of energy-metabolism related binding functions, such as NADP binding (Fig 1b).

Fig 1. Network visualization of Gene Ontology process and function annotation differences between normal water and drought treatments at the CA site. Significant Gene Ontology (GO) annotations from ENNB analysis were grouped by semantic similarity into a network. The size of each node is proportional to the frequency of annotation relative to the GO database. Similar terms are linked with edges. Circles and boxes indicate terms shared between field sites. a) CA field site GO Process annotations that were significantly different between fully watered and drought microbial phyllosphere samples. b) CA field site GO function annotations that were significantly different between fully watered and drought microbial phyllosphere samples.

Functional annotations that were significant from the Dumas-Etter, TX field site (abbreviated as DE) include a range of metabolic and regulatory terms, with a large cluster of amino acid, nucleic acid, and sugar metabolic enzymes (center of Fig 2a) and a second large cluster of regulatory and response categories (top of Figure 2a), such as quorum sensing. Topics related to response to oxidative stress form a smaller cluster. Unusual categories with single small nodes include protein refolding and reactive oxygen species metabolism. The term 'transcriptional regulation' was shared with the CA term list (circled in Fig 1 and Fig 2). The function term network (Fig 2b) also has a cluster for metal ion binding (visible at the top left of Fig 2b). After quality control, the DE site retained all six samples (Table 1) and this site had the largest number of significant

annotations (Fig 2, supplemental Fig S1).

Fig 2. Network visualization of Gene Ontology process and function annotation differences between normal water and drought treatments at the DE site. Significant Gene Ontology (GO) annotations from ENNB analysis were grouped by semantic similarity into a network. The size of each node is proportional to the frequency of annotation relative to the GO database. Similar terms are linked with edges. Circles and boxes indicate terms shared between field sites. a) DE field site GO Process annotations that were significantly different between fully watered and drought microbial phyllosphere samples. b) DE field site GO function annotations that were significantly different between fully watered and drought microbial phyllosphere samples.

Significant annotations from the Halfway, TX field site (abbreviated as HF) include a group of biosynthetic enzymes for amino and fatty acid synthesis (Fig 3a top left), and amino acid biosynthesis enzymes (Fig 3a top right). The process annotation 'translation' was shared with the DE site (indicated by the dashed square around the node and annotation label), and amino acid transport was shared with the CA field site (indicated by a dashed diamond). In the process listing, an example unusual pathway found only in HF is self proteolysis. Functional annotations include a set of regulatory activities (e.g., kinases) and several ion binding activities. The zinc ion binding activity was shared with the DE annotation list. One unusual annotation found only in HF function was cob(I)yrinic acid a,c-diamide adenosyltransferase, which is part of the vitamin B12 cofactor pathway.

Fig 3. Network visualization of Gene Ontology process and function annotation differences between normal water and drought treatments at the HF site. Significant Gene Ontology (GO) annotations from ENNB analysis were grouped by semantic similarity into a network. The size of each node is proportional to the frequency of annotation relative to the GO database. Similar terms are linked with edges. Circles and boxes indicate terms shared between field sites. a) HF field site GO Process annotations that were significantly different between fully watered and drought microbial phyllosphere samples. b) HF field site GO function annotations that were significantly different between fully watered and drought microbial phyllosphere samples.

Plant Traits

To confirm that drought treatment plots were relevant for host plant performance, we analyzed plant growth measurements. All plant measurements at all sites showed significantly less growth in the treatment with less water (Fig 4). Plot effects were examined for each trait and no significant interaction between plot and replicate was found (results not shown). Mid-season plant heights were significantly less ($P<0.0001$) in the drought condition for the CA site. The drought-treated plants were 20% shorter, with an estimated difference between normal water and drought of 0.158 meters. The DE field site with plot 101 excluded exhibited significant ($P=0.0139$) effects of drought on end-of-season seed weight (Fig 4b), with the seed weights in drought reduced by about 25% (estimated difference of 0.468 grams less in drought samples). Plot 101 from the 75% site had a late-season rain event after microbiome sampling but before seed harvest that necessitated its exclusion. Drought reduced seed weight by 50% at the HF field site (Fig 4c), with $P<0.0001$ and an effect difference of 1.206 g less in drought seed samples. Cob diameters were also significantly smaller in the drought-treated plants (Fig 4 d, e, f) with the effect size differences ranking the drought intensity of DE (1.66

mm less in drought) less than CA (2.52 mm less in drought), with the most severe cob diameter drought effects at the HF site (3.24 mm less in drought).

Fig 4. Drought effects on plant growth. Error bars are standard error and colors are grouped within a field site. a) Comparison of plant heights in drought and well-watered plots from the California-Albany (CA) field site; bar heights indicate average height in meters. Drought (less drip irrigation) n=40, well-watered regular drip irrigation n=40. b) Comparison of seed weights from mild drought (75% of normal irrigation) and well-watered (100% irrigation) field blocks at the Texas-Dumas-Etter (DE) field site. Colored bar indicates mean value. Drought n=18, well-watered n=17. c) Comparison of seed weights from in intense drought (half of normal watering level, DRT) and well-watered (WW) field plots at the Texas-Halfway (HF) site. Colored bar indicates mean value. DRT n=4 and WW n=22. Zeros (cobs with no seeds from DRT) were not included in the analysis. d) Box plot of cob diameter of CA samples; white line is mean and quantiles are indicated by the box and whiskers, n=12. e) Comparison of cob diameter by water treatment in samples from the DE site. Colored bar indicates mean value, n(75%)=28, n(100%)=26. f) Comparison of cob diameter by water treatment in samples from the HF site. Colored bar indicates mean value, n(DRT)=10, n(WW)=22.

Discussion

We qualitatively compared functional genes across all three sites (Supplemental Figure 1, synapse repository syn15575565), though we did not fit a statistical model for comparisons of drought effects across field sites as the field sites differed in multiple ways. There were more shared drought-treatment-relevant functional categories in comparisons of the CA and DE field sites than in comparisons with the HF site (Supplemental Figure 1). This suggests that drought severity could play a role in functional gene importance despite differences in soil and other aspects of each field environment, because the CA and DE plots did share milder drought conditions despite differences in delivery of irrigation. We would expect differences across field sites based on plant physiology and known differences in maize growth across field sites [44]. However, field site is confounded with the field-specific drought treatments in our study and we thus cannot quantitatively compare the field site annotation networks. Shared annotations across field sites often were not consistently increased or decreased in read count levels. For example, amino acid transport process read counts were higher in watered samples at the HF site and higher in drought samples at the CA site. However, the extent of drought-treatment significant annotation term sharing (without consideration of read count levels, as shown in Supplemental Figure 1) is consistent with the extent of plant growth effect, with HF sharing fewer terms and having more severe drought.

Lists of phyllosphere ribotypes from prior field studies [45–48] were used to generate a list of expected species. Expected phyllosphere species that were also present in our samples include Methylobacterium spp., Dietzia spp., and Pseudomonas spp., (Supplemental file metaphlan2.tsv in repository syn15575484). We carried out a detailed comparison of the annotations from the rice phyllosphere proteome [49] to our list. Six rice GO process were in the metaproteome pfam list [49], and three of the six were shared with our process lists. Recent literature on functional genes suggests that functions are more predictive than ribotype profiles [30, 31]. Therefore, testing the effects of synthetic communities with similar ribosomal but different functional composition would be of broad interest. Our functional gene information is a step

toward designing a future synthetic community test of functional annotation predictive ability.

In a maize leaf microbe association genetics experiment, predicted metabolic functions were more heritable than ribotypes, which also suggests that function is key [32]. Selection for specific microbial functional genes or generic markers for pathways could easily be incorporated into newer DNA-based crop genomic selection processes that are sequencing based [50–52]. The importance of incorporating microbial sequence predictors lends support to the movement toward sequencing to collect all DNA data, not just filtered SNP sets or SNPs with prior data on causality. Microbial sequences are not in linkage disequilibrium like chromosomal SNPs, so it would not be possible take advantage of tag SNPs. Because the cost of complete sequencing is decreasing, we advocate for modeling and tests of full-sequence predictors that include both host chromosomal and epiphyte functional DNA information.

We suggest that a key next step in understanding use of leaf microbial annotations for crop improvement would be to measure microbial community annotations in selected and unselected breeding program lines across multiple test sites. This would allow the estimation of the genotype and environment breeding values for functional gene annotation. That information would determine future breeding strategy and would be efficient, because collection of functional gene information could be an add-on to host breeding experiments such as g2f for maize (https://www.genomes2fields.org/) and terraRef for sorghum (http://terraref.org/). There are few publicly available field sites for drought experiments – we know of only five within the continental USA – so public-private partnerships and use of large-scale field experimental networks are logical next steps for better understanding of microbial community development for crop improvement.

Leaf epiphytes have short and long term intervention possibilities. Indirect selection for host effects is likely to be more cost-effective than inoculation, but that takes much longer to implement through the required multiple breeding cycles. Leaf microbes are typically not in seeds and thus not consumed. Thus, these microbes are logical targets for improved forage quality, energy extraction from biomass, or optimization of soil fertility for the next season as well as for plant host benefit.

We advocate for future experiments that build on the functional genes we identified and combining synthetic community development approaches with breeding experiments to generate knowledge that would be needed for future holobiont breeding system development. Our results allow prioritization of specific gene function pathways in choosing culturable microbe mixtures for future experiments on design of drought tolerant epiphytic communities.

Conclusion

We identified a range of biosynthetic and regulatory microbial functional and process annotations that differed between drought and well-watered maize leaf epiphytic communities. These functions now provide a target for selection of beneficial microbes and for design of synthetic community casual tests of community interactions.

Acknowledgments

We thank Neha Gupta, Bryan Frank and Kelvin Li for their work on library construction and sequencing. We are obliged to Stephen P. Talley for his hard work on the ENNB analysis of an initial annotation dataset. We very much appreciate the contributions of Sarah Hake, China Lunde and the Hake lab members, who provided

field and lab space for this project and carried out the field management for us. We are grateful to Robert L. Bryden, Danielle Allery Nail, and Bonnie M. Mitchell for assistance with plant trait measurements and to Monika Bihan for assistance with the sequencing and annotations. Author Roles: Ann E. Stapleton conceived the experiment, designed and carried out the field sampling, oversaw the data analysis, and wrote the manuscript. Wenwei Xu set up the Texas field sites with drought and normal irrigation treatments, planted the experimental plots, collected trait data, and edited the manuscript. Stephen P. Talley analyzed the annotation count data. Kelvin Li analyzed the taxonomic data for the 13 sample subset. Stuart Gordon and Brad Goodner edited the manuscript. Barbara Methe supervised the sequencing and sequence data processing and edited the manuscript.

References

1. Bulgarelli D, Schlaeppi K, Spaepen S, van Themaat EVL, Schulze-Lefert P. Structure and Functions of the Bacterial Microbiota of Plants. Annual Review of Plant Biology. 2013;64(1):807–838. doi:10.1146/annurev-arplant-050312-120106.

2. Müller DB, Vogel C, Bai Y, Vorholt JA. type [; 2016]Available from: `http://www.annualreviews.org/doi/10.1146/annurev-genet-120215-034952`.

3. Berg G, Grube M, Schloter M, Smalla K. Unraveling the plant microbiome: looking back and future perspectives. Plant Biotic Interactions. 2014;5:148. doi:10.3389/fmicb.2014.00148.

4. Knief C. Analysis of plant microbe interactions in the era of next generation sequencing technologies. Plant Genetics and Genomics. 2014;5:216. doi:10.3389/fpls.2014.00216.

5. Mine A, Sato M, Tsuda K. Toward a systems understanding of plant–microbe interactions. Frontiers in Plant Science. 2014;5. doi:10.3389/fpls.2014.00423.

6. Morella NM, Weng FCH, Joubert PM, Metcalf CJE, Lindow S, Koskella B. Successive passaging of a plant-associated microbiome reveals robust habitat and host genotype-dependent selection. bioRxiv. 2019; p. 627794. doi:10.1101/627794.

7. Schlechter RO, Miebach M, Remus-Emsermann MNP. Driving factors of epiphytic bacterial communities: A review. Journal of Advanced Research. 2019;doi:10.1016/j.jare.2019.03.003.

8. Vacher C, Hampe A, Porté AJ, Sauer U, Compant S, Morris CE. The Phyllosphere: Microbial Jungle at the Plant–Climate Interface. Annual Review of Ecology, Evolution, and Systematics. 2016;47(1):1–24. doi:10.1146/annurev-ecolsys-121415-032238.

9. Leach JE, Triplett LR, Argueso CT, Trivedi P. Communication in the Phytobiome. Cell. 2017;169(4):587–596. doi:10.1016/j.cell.2017.04.025.

10. Howden AJM, Preston GM. Nitrilase enzymes and their role in plant–microbe interactions. Microbial biotechnology. 2009;2(4):441–451. doi:10.1111/j.1751-7915.2009.00111.x.

11. Choudhary DK, Sharma KP, Gaur RK. Biotechnological perspectives of microbes in agro-ecosystems. Biotechnology Letters. 2011;33(10):1905–1910. doi:10.1007/s10529-011-0662-0.

12. Churchland C, Grayston SJ. Specificity of plant-microbe interactions in the tree mycorrhizosphere biome and consequences for soil C cycling. Frontiers in Microbiology. 2014;5:261. doi:10.3389/fmicb.2014.00261.

13. Phieler R, Voit A, Kothe E. Microbially supported phytoremediation of heavy metal contaminated soils: strategies and applications. Advances in Biochemical Engineering/Biotechnology. 2014;141:211–235.

14. Orozco-Mosqueda MdC, Rocha-Granados MdC, Glick BR, Santoyo G. Microbiome engineering to improve biocontrol and plant growth-promoting mechanisms. Microbiological Research. 2018;208:25–31. doi:10.1016/j.micres.2018.01.005.

15. Parnell JJ, Berka R, Young HA, Sturino JM, Kang Y, Barnhart DM, et al. From the Lab to the Farm: An Industrial Perspective of Plant Beneficial Microorganisms. Frontiers in Plant Science. 2016;7. doi:10.3389/fpls.2016.01110.

16. Mitter B, Pfaffenbichler N, Flavell R, Compant S, Antonielli L, Petric A, et al. A New Approach to Modify Plant Microbiomes and Traits by Introducing Beneficial Bacteria at Flowering into Progeny Seeds. Frontiers in Microbiology. 2017;8. doi:10.3389/fmicb.2017.00011.

17. Vorholt JA, Vogel C, Carlström CI, Müller DB. Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research. Cell Host & Microbe. 2017;22(2):142–155. doi:10.1016/j.chom.2017.07.004.

18. Bringel F, Couée I. Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics. Frontiers in Microbiology. 2015;6. doi:10.3389/fmicb.2015.00486.

19. Remus[U+2010]Emsermann MNP, Schlechter RO. Phyllosphere microbiology: at the interface between microbial individuals and the plant host. New Phytologist. 2018;218(4):1327–1333. doi:10.1111/nph.15054.

20. Müller C, Riederer M. Plant surface properties in chemical ecology. Journal of Chemical Ecology. 2005;31(11):2621–2651. doi:10.1007/s10886-005-7617-7.

21. Luziatelli F, Ficca AG, Colla G, Baldassarre Švecová E, Ruzzi M. Foliar Application of Vegetal-Derived Bioactive Compounds Stimulates the Growth of Beneficial Bacteria and Enhances Microbiome Biodiversity in Lettuce. Frontiers in Plant Science. 2019;10. doi:10.3389/fpls.2019.00060.

22. Kerdraon L, Laval V, Suffert F. How can a knowledge of microbiota-pathogen interactions in cereal cropping systems help us to manage residue-borne fungal diseases? arXiv:190302246 [q-bio]. 2019;.

23. Compant S, Samad A, Faist H, Sessitsch A. A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. Journal of Advanced Research. 2019;doi:10.1016/j.jare.2019.03.004.

24. Schmidhuber J, Tubiello FN. Global food security under climate change. Proceedings of the National Academy of Sciences. 2007;104(50):19703–19708. doi:10.1073/pnas.0701976104.

25. Kell DB, Kaprelyants AS, Weichart DH, Harwood CR, Barer MR. Viability and activity in readily culturable bacteria: a review and discussion of the practical issues. Antonie Van Leeuwenhoek. 1998;73(2):169–187.

26. Torsvik V, Øvreås L. Microbial diversity and function in soil: from genes to ecosystems. Current Opinion in Microbiology. 2002;5(3):240–245.

27. Rajendhran J, Gunasekaran P. Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. Microbiological Research. 2011;166(2):99–110. doi:10.1016/j.micres.2010.02.003.

28. Methé BA, Lasa I. Microbiology in the 'omics era: from the study of single cells to communities and beyond. Current Opinion in Microbiology. 2013;16(5):602–604. doi:10.1016/j.mib.2013.10.002.

29. Kelly LW, Williams GJ, Barott KL, Carlson CA, Dinsdale EA, Edwards RA, et al. Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors. Proceedings of the National Academy of Sciences. 2014;111(28):10227–10232. doi:10.1073/pnas.1403319111.

30. Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T. Bacterial community assembly based on functional genes rather than species. Proceedings of the National Academy of Sciences. 2011;108(34):14288–14293. doi:10.1073/pnas.1101591108.

31. Doolittle WF, Inkpen SA. Processes and patterns of interaction as units of selection: An introduction to ITSNTS thinking. Proceedings of the National Academy of Sciences. 2018;115(16):4006–4014. doi:10.1073/pnas.1722232115.

32. Wallace JG, Kremling KA, Buckler ES. Quantitative Genetic Analysis of the Maize Leaf Microbiome. bioRxiv. 2018; p. 268532. doi:10.1101/268532.

33. Sen TZ, Andorf CM, Schaeffer ML, Harper LC, Sparks ME, Duvick J, et al. MaizeGDB becomes 'sequence-centric'. Database. 2009;2009:bap020. doi:10.1093/database/bap020.

34. Langmead B. A tandem simulation framework for predicting mapping quality. Genome Biology. 2017;18:152. doi:10.1186/s13059-017-1290-3.

35. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, et al. Improved maize reference genome with single-molecule technologies. Nature. 2017;546(7659):524–527. doi:10.1038/nature22971.

36. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016;4:e2584. doi:10.7717/peerj.2584.

37. Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL, et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS computational biology. 2012;8(6):e1002358. doi:10.1371/journal.pcbi.1002358.

38. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nature Methods. 2015;12(10):902–903. doi:10.1038/nmeth.3589.

39. Pookhao N, Sohn MB, Li Q, Jenkins I, Du R, Jiang H, et al. A two-stage statistical procedure for feature selection and comparison in functional analysis of metagenomes. Bioinformatics. 2015;31(2):158–165. doi:10.1093/bioinformatics/btu635.

40. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS ONE. 2011;6(7):e21800. doi:10.1371/journal.pone.0021800.

41. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13(11):2498–2504. doi:10.1101/gr.1239303.

42. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Research. 2005;33(Web Server issue):W741–748. doi:10.1093/nar/gki475.

43. Mee MT, Collins JJ, Church GM, Wang HH. Syntrophic exchange in synthetic microbial communities. Proceedings of the National Academy of Sciences. 2014; p. 201405641. doi:10.1073/pnas.1405641111.

44. Carena MJ, Hallauer AR, Miranda Filho JB, Filho JBM. Quantitative Genetics in Maize Breeding. Springer New York; 2010.

45. Rastogi G, Sbodio A, Tech JJ, Suslow TV, Coaker GL, Leveau JHJ. Leaf microbiota in an agroecosystem: spatiotemporal variation in bacterial community composition on field-grown lettuce. The ISME journal. 2012;6(10):1812–1822. doi:10.1038/ismej.2012.32.

46. Copeland JK, Yuan L, Layeghifard M, Wang PW, Guttman DS. Seasonal Community Succession of the Phyllosphere Microbiome. Molecular Plant-Microbe Interactions. 2015;28(3):274–285. doi:10.1094/MPMI-10-14-0331-FI.

47. Wagner MR, Lundberg DS, Rio TGd, Tringe SG, Dangl JL, Mitchell-Olds T. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. Nature Communications. 2016;7:12151. doi:10.1038/ncomms12151.

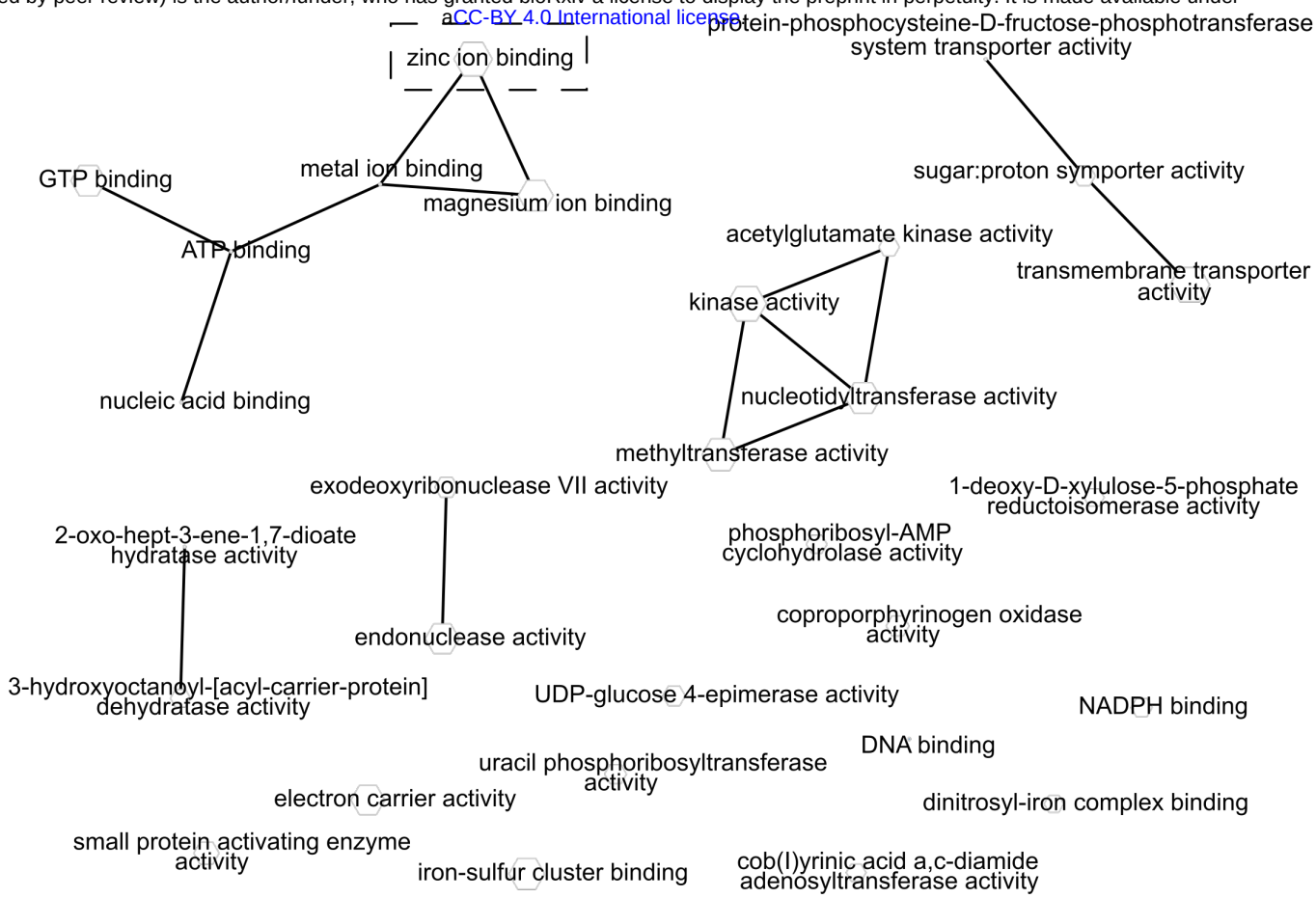
48. Bodenhausen N, Horton MW, Bergelson J. Bacterial Communities Associated with the Leaves and the Roots of Arabidopsis thaliana. PLOS ONE. 2013;8(2):e56329. doi:10.1371/journal.pone.0056329.

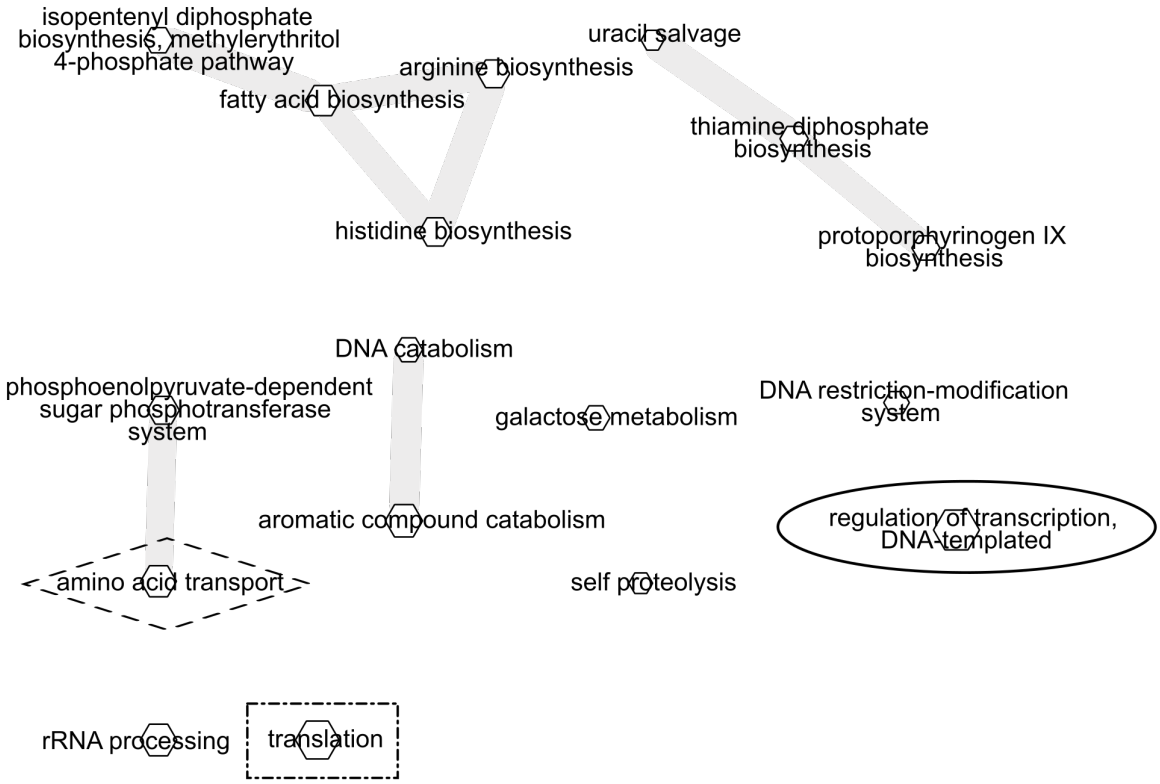
49. Knief C, Delmotte N, Chaffron S, Stark M, Innerebner G, Wassmann R, et al. Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. The ISME Journal. 2012;6(7):1378–1390. doi:10.1038/ismej.2011.192.

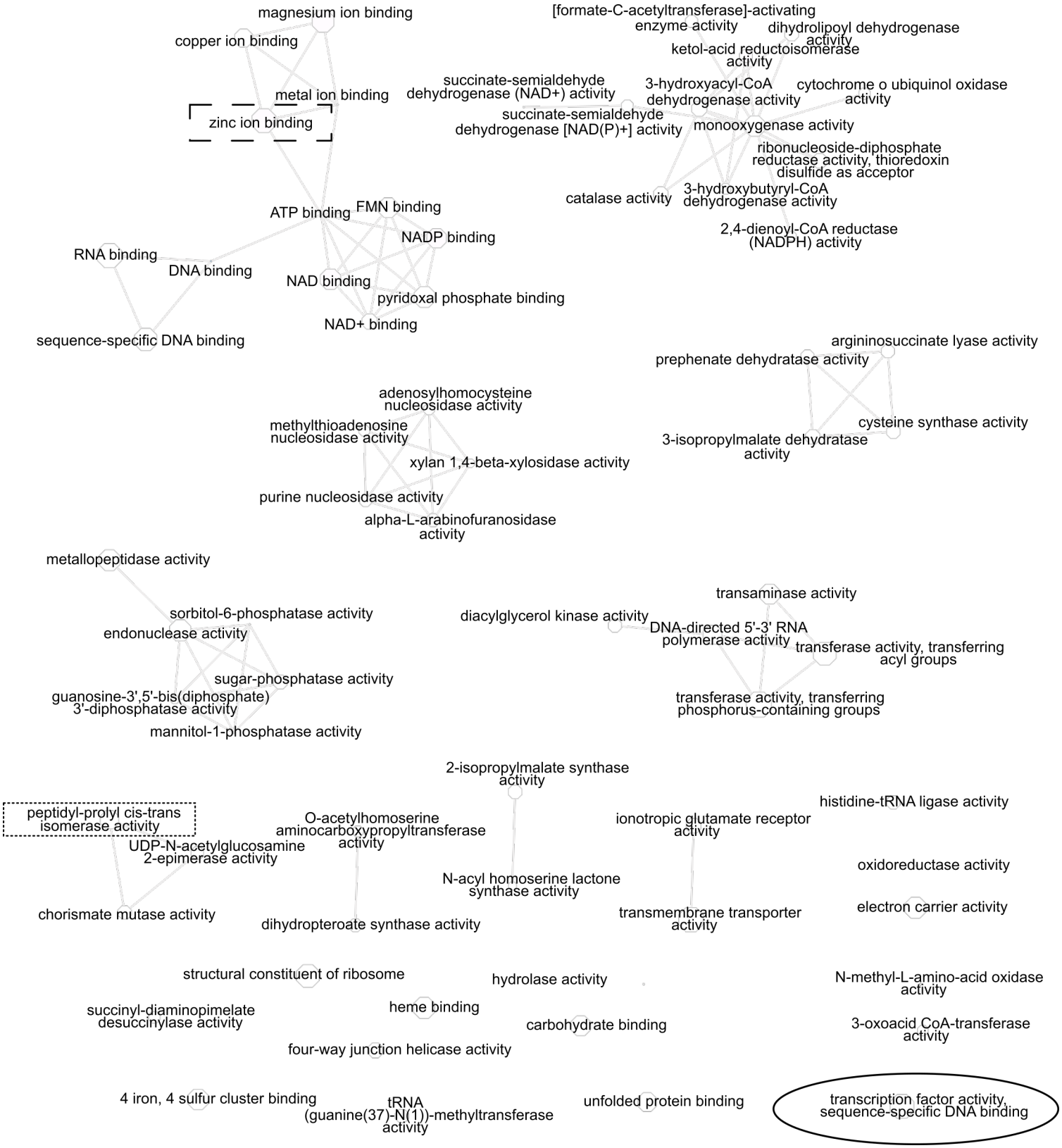
50. Pérez-Enciso M, Forneris N, Campos Gdl, Legarra A. Evaluating Sequence-Based Genomic Prediction with an Efficient New Simulator. Genetics. 2016; p. genetics.116.194878. doi:10.1534/genetics.116.194878.

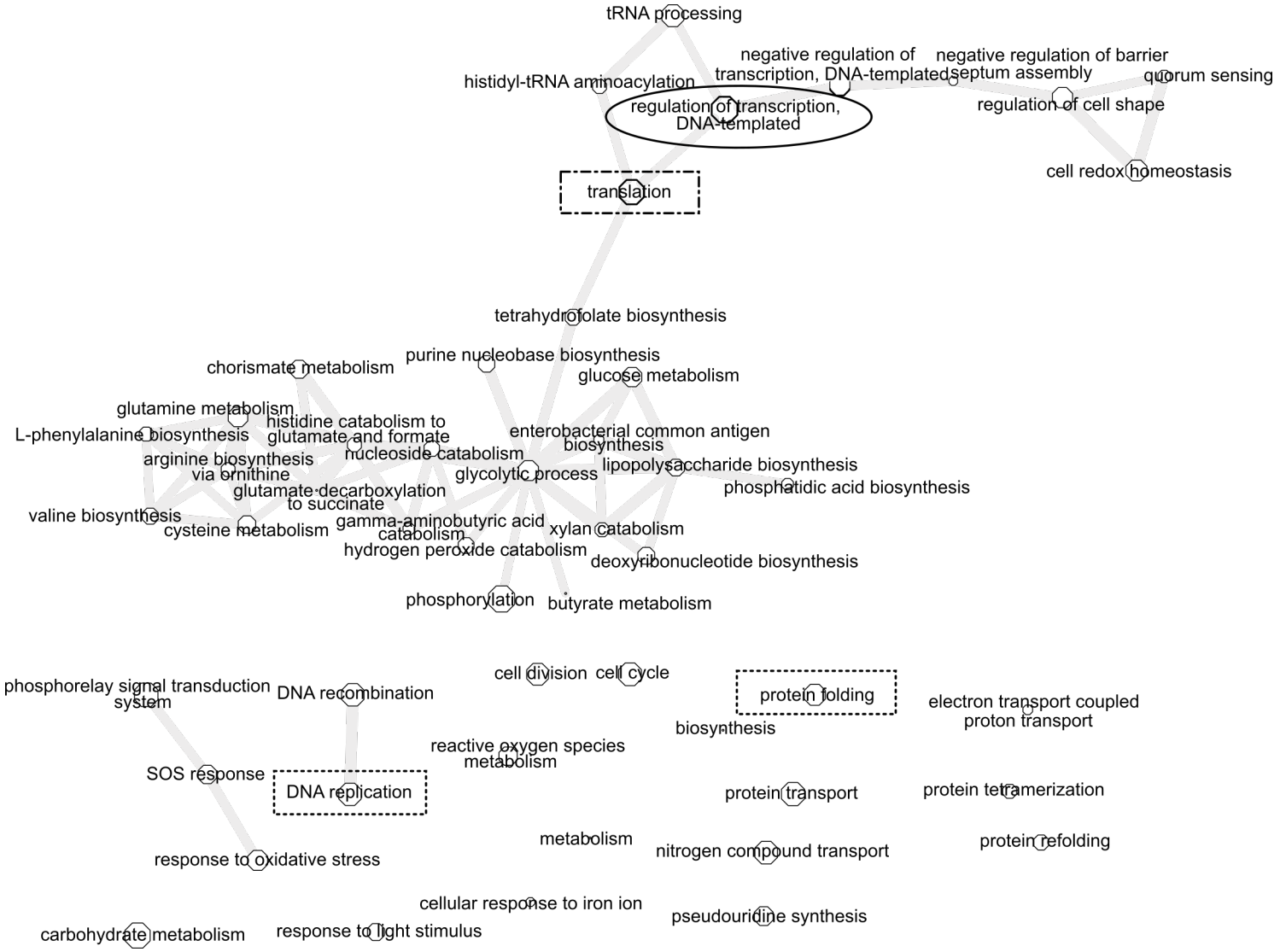
51. Forneris NS, Vitezica ZG, Legarra A, Pérez-Enciso M. Influence of epistasis on response to genomic selection using complete sequence data. Genetics Selection Evolution. 2017;49:66. doi:10.1186/s12711-017-0340-3.

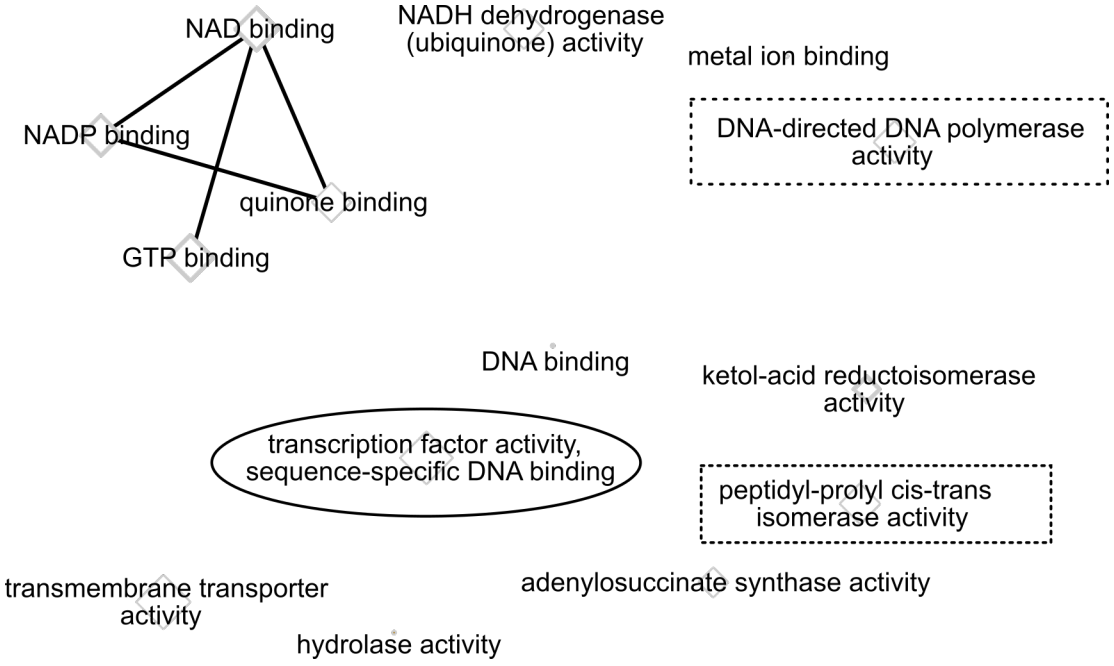
52. Zhang C, Kemp RA, Stothard P, Wang Z, Boddicker N, Krivushin K, et al. Genomic evaluation of feed efficiency component traits in Duroc pigs using 80K, 650K and whole-genome sequence variants. Genetics Selection Evolution. 2018;50:14. doi:10.1186/s12711-018-0387-9.

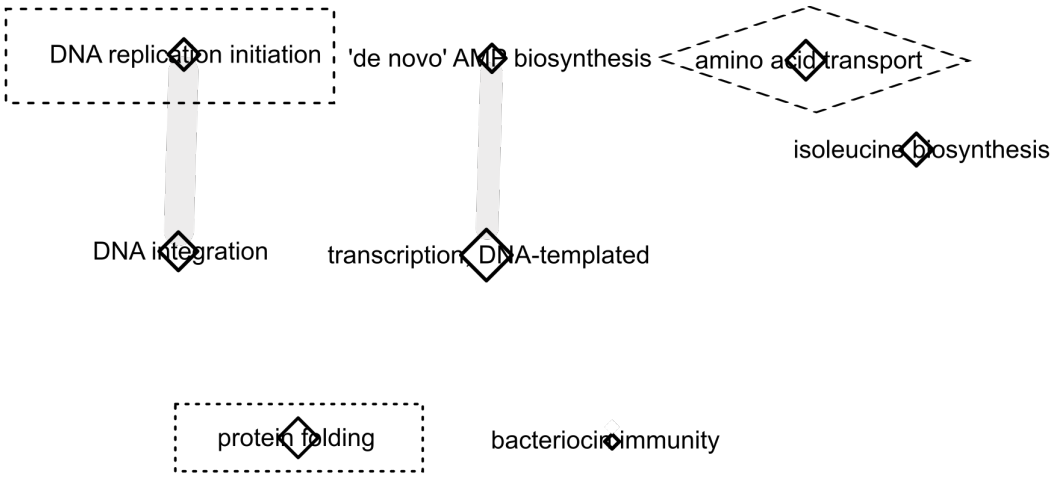


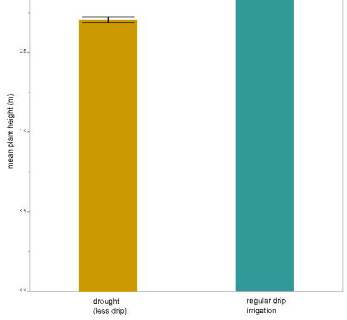


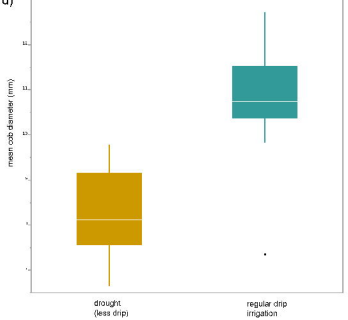


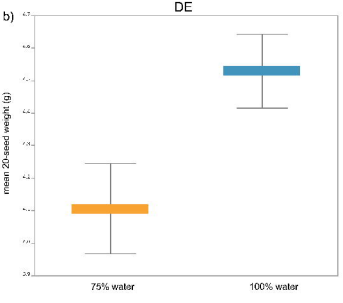


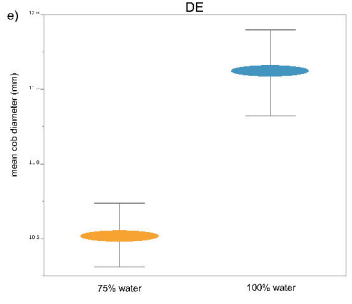


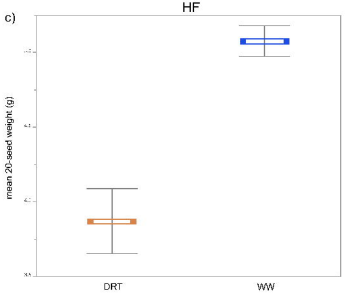


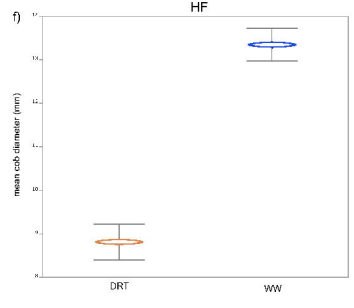

mean plant height (m)
drought
(less drip)
regular drip
irrigation


d)
mean cob diameter (mm)
drought
(less drip)
regular drip
irrigation


b)
DE
mean 20-seed weight (g)
75% water
100% water


e)
DE
mean cob diameter (mm)
75% water
100% water


c)
HF
mean 20-seed weight (g)
DRT
WW


f)
HF
mean cob diameter (mm)
DRT
WW