UNIVERSITY OF NEBRASKA AT OMAHA

Life in the Phyllobiome: Functional Adaptations in *Novosphingobium sp.* 'Leaf2', a Leaf-Borne Alphaproteobacteria

Katie Sindelar | Dr. Dhundy Bastola
Department of Bioinformatics, University of Nebraska at Omaha, Omaha, NE 68182

ABSTRACT

Plant-associated microbiomes have emerged as a significant influence on host health and development, driving interest into the functional repertoires of constituent organisms and the mechanisms of host selection on resident microbial populations. Research into plant commensal bacteria have largely focused on rhizospheric milieus - the leaf-surface phyllobiome presents a more punishing environment, where microbia are subjected to high levels of UV radiation, low water and nutrient availability, and foliar agricultural chemicals in food crops. To investigate adaptations towards success in this harsh environment, a comparative genomics analysis across a cohort of *Novosphingobium* species was conducted using public bioinformatics resources and tools. This research identifies several features expected to contribute to dominance in the phyllobiome for *Novosphingobium. sp.* 'Leaf2', including a novel type IV secretion system, enzymes to degrade plant compounds, and three LuxR 'solo' genes which enable the bacterium to respond to plant small molecules, typically AHLs (N-acyl homoserine lactones). Insight into these functional adaptations shed light on the mechanisms of mutualistic interspecies relationships and may lead to new means and biological agents to promote food safety and plant health.

BACKGROUND

In a 2015 report, Ceuppens et. al. investigated **epiphytic bacteria on commercial basil leaves, identifying *Novosphingobium* as the predominant genus present**[1] – notably, spoilage rate was not associated with any bacterial group. As basil leaves contain a range of volatile aromatic compounds with potent antibacterial activities, the dominance of *Novosphingobium* in the phyllobiome suggests niche co-selection by both host and commensal as well as a biochemical skillset in the latter for countering plant chemical defenses. ***Novosphingobium* is a highly versatile genus of alphaproteobacteria**, including members associated with contaminated soil, fresh, marine, and mine waste water, tissues, and in one case human immune disease[2]. These bacteria are characterized by **enzymes to degrade aromatic compounds** such as pesticides and persistent organic pollutants, as well as **unique cell wall glycosphingolipids** which may play a role in evading host immune responses. To characterize specific adaptations towards life in the phyllobiome, a comparative genomics approach was conducted with a focus on *Novosphingobium sp.* 'Leaf2' (LEAF), a publicly available strain retrieved from *A. thaliana*.

METHODOLOGY

A cohort of 8 representative strains of *Novosphingobium* was chosen, including four plant-associated bacteria and four environmental strains, and genomes downloaded from NCBI. Phylogenetic relationships between cohort members were assessed using MEGA. **Pangenome analysis** for this cohort was performed using BPGA in order to assess canonical and divergent features for this subset. **Biological subsystem and biochemical pathway participation** across the cohort was provided by PATRIC[3]. PATRIC was also used to assign **intergenus protein family annotations** to cohort members and select out features unique to LEAF. IslandViewer4 was used to predict **genome islands, sites of acquired symbiosis/virulence genes**. A 'host' transcriptome for sweet basil (*Ocimum basilicum)* was assembled and a BLAST database of **LEAF features were matched against plant-expressed sequences** in search of conjugated xenologs or incidental homologs benefiting molecular crosstalk.

COHORT MEMBERS

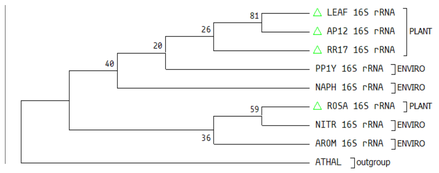
	CODE	GENOME ID	SPECIES NAME
PLANT	LEAF	1735670.3	***Novosphingobium sp.* Leaf2**
	AP12	1144305.3	*Novosphingobium sp.* AP12
	ROSA	1219053.3	*Novosphingobium rosa* NBRC 15208
	RR17	555793.3	*Novosphingobium sp.* Rr 2-17
ENVIRO	AROM	279238.18	*Novosphingobium aromaticivorans* DSM 12444
	NAPH	1219034.3	*Novosphingobium naphthalenivorans* NBRC 102051
	NITR	983920.3	*Novosphingobium nitrogenifigens* DSM 19370
	PP1Y	702113.7	*Novosphingobium sp.* PP1Y

RESULTS

COHORT PHYLOGENY

Fig. 1: MEGA phylogenetic tree output. Branch numbers show percentage of replicate trees containing the indicated cluster (bootstrap replicates n=500). Maximum Likelihood method based on the Tamura-Nei model [1]. ATHAL: *A. thaliana* 16S rRNA (from chloroplast) used as outgroup. Green triangles mark plant-associated cohort members.

MEGA: Kumar, S., Stecher, G., and Tamura, K. "MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets." (2016) Molecular Biology and Evolution (2016). 33:1870-1874



PANGENOME KEGG PATHWAY ANALYSIS

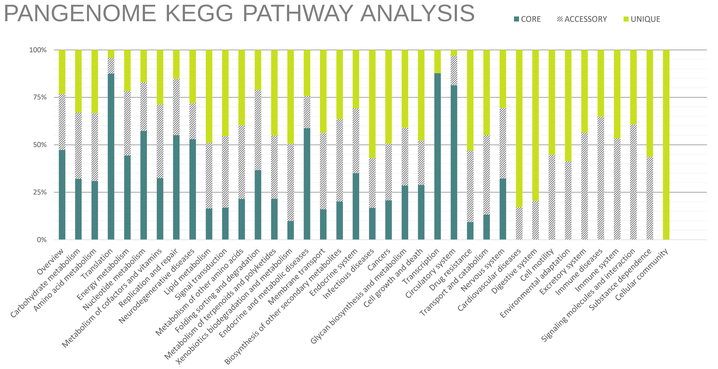


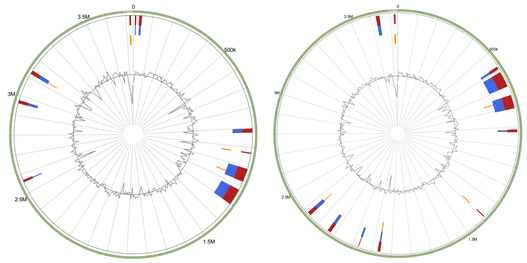
Fig. 2: BPGA pangenome pathway map output. Shows proportion of each KEGG pathway divided into core, accessory, and unique subsets of the NOVO cohort pangenome, with pathways in order of decreasing proportion of the overall core genome. Though this characterization of the *Novosphingobium* pangenome reveals trends in biochemical categories where strains are more likely to specialize, it excludes all hypothetical proteins without annotated functions, which represent 56.4% of cohort proteins (10,926 of 19,365 intergenus protein families). Until these protein products can be validated and characterized, pathway analysis results are likely not accurate representations of all functional biochemical activities.

BPGA: Chaudhari, Narendrakumar M., Vinod Kumar Gupta, and Chitra Dutta. "BPGA-an ultra-fast pan-genome analysis pipeline." Scientific reports 6, 24373 (2016). doi: 10.1038/srep24373.

LEAF GENOME ISLAND ANALYSIS

Fig. 3: IslandViewer4 output for LEAF genome anchored to AROM and PP1Y genomes. Central plotline indicates local GC distribution. Color of identified regions indicates prediction method used by the tool: Integrated, IslandPath-DIMOB, or SIGI-HMM. The predictions were in consensus on 222 features.

IslandViewer4: Bertelli, C. et al. 2017. "IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets" Nucleic Acids Research. 2017 May 2. pathogenomics.sfu.ca/islandviewer/



LEAF vs AROM
***260** features identified*

LEAF vs PP1Y
***250** features identified*

FEATURE SETS

UNIQUE TO COHORT	UNIQUE TO LEAF	GENOME ISLANDS	HITS TO BASIL TRANSCRIPTOME
37484	969	220	136

Fig. 4: In order to locate the most **probable** niche-adaptive features in LEAF, features shared with any other cohort members were excluded from targeted analysis. This reduced set was subsequently intersected with the 2 specialty feature lists, from which annotated genes of interest were selected for discussion.

DISCUSSION

BIOCHEMICAL FEATURES

Of the features unique to LEAF, 83.2% (806) were hypothetical, limiting higher-level analysis. Among annotated proteins, 39 could be assigned KEGG pathways, several of which were associated with enzymes for plant compounds. **Fatty acid metabolism represented 21% of unique pathway features.** This category was highly variable in the pangenome, indicating niche specialization. Genes for degradation of limonene/pinene (5) and geraniol (2) were also noted. Biosynthesis genes for flavonoid (2), stilbenoid/diarylheptanoid/gingerol (2), and carotenoid (2) suggest that this bacterium may be involved in diversification of host plant secondary metabolites. Across the cohort, LEAF exhibited the highest numbers of enzymes for vanillate, syringate, and gallic acid, which were found in all but 2 ENVIRO genomes.

TYPE IV SECRETION SYSTEM SIMILAR TO TI PLASMID

Genome island analysis identified a contiguous set of VirB/VirD genes highly similar to those found on the *A. tumefaciens* Ti pathogenicity plasmid**. In *A. tumefaciens,* this operon allows the bacterium to conjugate plasmid-encoded DNA into a host plant's genome and elicit the production of a novel metabolite** on which it can subsist, ultimately leading to host tumorigenesis. Intriguingly, one component (VirB10) was also identified in the assembled basil transcriptome, suggesting that the source plant material may have been compromised by a similar organism at the time of sequencing. These host control mechanisms feature species-specific effector molecules to manipulate host gene expression. While T4SS genes were found in all but 2 cohort genomes (AP12 and NITR), only those found in LEAF were annotated as the Ti-type variants, which may indicate independent acquisition from community microbia.

LUXR SOLO GENES

LEAF contained 3 unique LuxR solos. These genes belong to a family of AHL-dependent quorum-sending proteins, but **in plant-associated bacteria, these bifunctional proteins are known to interact with plant signalling compounds and initiate gene expression**[4,5]. In one experiment, 3 LuxR-like solos identified in an endophytic, pathogenic strain of *Pseudomonas syringae* revealed that *in planta* survival was compromised when these genes were mutated, and that their effect on survival was additive[6]. LuxR responds to AHLs produced by LuxI synthases - these synthases were found in all PLANT genomes as well as 2 ENVIRO cohort members. We expect that these proteins are critical to benign survival within the phyllobiome, though they may retain the quorum-sensing capabilities observed in other species to provide bacterial population sensing. Further experimentation is needed.

CONCLUSIONS

While we often consider bacteria masters of competitive niche exploitation, plant microbiomes present numerous examples of promiscuous gene sharing and conjugation among unrelated microbiome members, allowing for strain-level specialization and local population-level competence within a host site. Though specific to LEAF, the features identified in this *in silico* exploration echo findings in divergent plant-associated bacteria, and present opportunities for further research into the nature of these complex, cooperative consortia.

ACKNOWLEDGEMENTS

This research was undertaken with the support of a Funding for Undergraduate Scholarly Research (FUSE) grant and the advisement of Dr. Dhundy Bastola of the UNO Bioinformatics Department. In order to preserve space, references will be made available separately on the date of presentation.

University of Nebraska Omaha