

## **Priorities for the *Xanthomonas* research community**

A white paper summarizing discussions held on July 15, 2009 at the *Xanthomonas* Genomics Conference at Pingree Park, Colorado.

Approximately 65 researchers gathered for *Xanthomonas* Genomics Conference 2009 from July 13 to 15 at Pingree Park, Colorado. Topics presented in talks and posters ranged from comparative and functional genomics to gene regulation and signaling in host-pathogen interactions to high throughput sequencing and data management. On the last day, attendees participated in roundtable discussions to identify challenges and opportunities, priorities for research, and resources that will enable the community to rapidly advance useful understanding of this important group of plant pathogenic bacteria.

In this paper, we present consensus views that resulted from these discussions for three topics: 1) type III effector nomenclature, 2) targets for additional genome sequencing, and 3) platforms for sharing information, data, and materials. The purpose of the paper is to spur and guide community initiatives, to inform research funding agencies and other interested parties, and to provide a benchmark to measure progress against at the next *Xanthomonas* Genomics conference, scheduled for March 2012.

### **Type III effector nomenclature**

Genomic and functional analyses of *Xanthomonas* species have uncovered an expanding number of candidate *Xanthomonas* effectors that are substrates for the type III secretion system (T3SS). Consensus centered on the need for an organized Xop database and consistent guidelines for Xop nomenclature, although it was generally agreed that any ontology used in the database should accommodate nomenclature of genes established in the literature.

Researchers are encouraged to categorize and name new effectors using an agreed upon systematic nomenclature, and where appropriate, note proper categorization for previously identified effectors for which the established name does not adequately reflect the category. It was noted that this approach would aid researchers within and outside the field both in conceptual understanding of effector function and in practical aspects of communication regarding effectors. This proposal is similar in scope to the system for Hop effectors of *Pseudomonas syringae* and related strains, which can be accessed at [pseudomonas-syringae.org](http://pseudomonas-syringae.org).

Regardless of the specific nomenclature adopted, the group agreed on the need for the *Xanthomonas* community to:

1. Establish and maintain an "official" Xop effector database on the world-wide web, curated by an individual laboratory.
2. Establish a database committee and staff (voluntary and/or grant supported) to review community submissions, annotate, and update information.
3. Adopt a robust nomenclature that follows conventions for bacterial genes and conveys the source of the gene.

4. Establish an electronic community forum (chat room, listserv, or the like) for comments and dissemination of information.
5. Apply stringent phylogenetic classification of strains and effector sequences in naming and adding candidate Xop effectors to the database.
6. Provide all available information regarding validation of type III delivery, function and structure, homology, and references for each Xop effector to be listed in the database.

With regard to specific nomenclature guidelines and database organization, the group did not reach a consensus, and all recognized the difficulties in designing a perfect system. The difficulties arise from the arbitrariness of strain names, the inconsistency of historical gene names and the dichotomy of avirulence and virulence effector functions reflected in historical names, mercurial phylogenetic assessments of strains, and changing species and pathovar designations. Another challenge is deciding on a basis for effector grouping. At present, most grouping is based on sequence relatedness and not necessarily function. But important questions persist. What level of relatedness justifies inclusion in the same group? Should the list include subgroups? And how should apparent chimeric effectors be classified?

A surprisingly difficult question was whether the community should adopt the Xop (*Xanthomonas* outer protein) nomenclature, modeled after the *Yersinia* (Yop) and *Pseudomonas syringae* (Hop) conventions. If the Xop nomenclature is accepted, questions arise regarding traditional names and sub-classifications. For example:

1. What Xop names are to be given, if any, to founding genes, whose names are inexorably linked to the literature and general usage?
2. Should AvrBs1, AvrBs2, and AvrBs3, specifically, be assigned to a Xop group, and if so, should they be designated XopA –XopC to reflect their priority, even though it would require renaming the present XopA, XopB and XopC?
3. And, should an Xop subscript be assigned that conveys strain or other specific information or would an appended number suffice, linking to necessary phylogenetic or source information in the official database?

Though details remain to be worked out, a majority of attendees voted to maintain the founding Avr and Xop names established in the literature while adopting a universal Xop nomenclature for candidate effectors annotated in current and future genome projects.

### **Targets for additional genome sequencing**

Genomes from most important bacterial pathogens of humans, animals and plants have now been sequenced, as have genomes from many commensal, symbiotic, and environmental microorganisms (Pallen and Wren, 2007). Analysis and comparison of these sequences have uncovered unexpected aspects of pathogen biology, including the important roles of horizontal gene transfer and genome decay in pathogen evolution. Most pathogens have evolved means to actively manipulate their host organism, among them the type III and type IV secretion systems, which deliver effector proteins and nucleic acids directly into host cells. For many of the effector proteins, a role in suppressing the basal defense response of plant hosts has been demonstrated (Guo et al., 2009). Surprisingly, homologues or remnants of type III secretion systems are also found in commensal strains and in environmental bacteria. Presumably, these

systems had a role in a former niche. Or, they may mediate yet uncharacterized interactions with nematodes, insects, or even single celled eukaryotes.

To fully understand the evolution and adaptation of plant pathogens, including *Xanthomonas* species, it is important to elucidate the repertoire of type III effectors and other pathogenicity factors at a genomic scale. Such knowledge will help to develop strategies to combat these bacteria and to breed durably resistant crops.

Ten complete high-quality genome sequences of *Xanthomonas* have been released. Selection of these strains was guided by their being well-characterized, well-studied organisms with value as models. Nevertheless, they represent only four of the 27 known species of *Xanthomonas*, and our picture of the genus as a whole is still fragmentary. With the recent progress in DNA sequencing technology and the dramatic reductions in cost, some comparative genomics projects have been initiated, yet these still focus on only a few selected *Xanthomonas* species and pathovars. To acquire a global view of pathogenicity and to better understand the evolution and distinguishing traits of the diversity within *Xanthomonas*, a more systematic approach is needed, both with respect to targets and techniques.

*Targets.* During the conference, strong interest was expressed a) to sequence non-pathogenic isolates of *Xanthomonas* and b) to fill the gaps in species-level representation. To obtain a broader view of *Xanthomonas* ecosystems, metagenomic projects should be initiated, focusing for example on epiphytic communities, xylem sap of diseased plants, and samples from irrigated fields. To better understand diversity and evolutionary relationships within the genus, representative strains of the currently unrepresented species should be sequenced.

*Techniques.* Initial projects using next-generation DNA sequencing technology (MacLean et al., 2009) have shown that a combination of deep-coverage 454 (~25 x coverage) and Solexa/Illumina (~50 x coverage) sequencing can result in draft genome sequences with most of the coding sequences, including those for type III effectors, completely revealed (Aury et al., 2008; Almeida et al., 2009; Reinhardt et al., 2009). Such high-quality draft genome sequences will suffice to tackle many key questions in bacterial phytopathology. However, high-copy number gene families, such as transposase genes of IS elements, and repetitive sequences, such as those in CRISPR loci and TAL effector genes, pose problems for the full assembly of genome sequences using these methods. These might be resolved by bioinformatic and technological advances. Yet for the time being, to obtain these sequences and to understand evolution of genome structure, a full assembly based on supplementary DNA sequencing by Sanger (or Pacific Bioscience, upon commercial launch) will be required. Alternatively, high-throughput next-generation sequencing of cosmid libraries, using barcodes, might be applied. A fully assembled reference genome for each *Xanthomonas* species would help in assembly and annotation of further strains and is therefore highly desirable.

With respect to annotation, participants noted that community efforts, such as used for *X. oryzae* pv. *oryzae* PXO99<sup>A</sup>, could ensure high efficiency and reliability (Salzberg et al., 2008). It was also noted that a combination of 454 sequencing of genomic DNA and Solexa/Illumina sequencing of bar-coded cDNA from bacteria cultured under diverse conditions would improve genome annotation. With respect to transcriptomics, a majority agreed that given the commercial availability of custom DNA chips and the emergence of methods for transcript profiling through deep sequencing, there is little need for construction of new spotted oligonucleotide or PCR fragment microarrays.

In opposition to the majority opinion that sequencing a representative of each of the 27 *Xanthomonas* spp. is a priority, some participants prioritized sequencing a few models in more depth. This is the current trend of some research teams in the United States and France, who proposed to sequence several strains of *X. oryzae* pv. *oryzae*, *X. citri* pv. *citri* and *X. axonopodis* pv. *phaseoli*. Despite the differences in priority, however, all agreed that, ultimately, sequencing of both taxonomically broad and taxonomically deep collections is desirable.

### **Platforms for sharing information, data, and materials**

With advances in genome sequencing have come genomic-based research resources that include full-length cDNA sequences, microarrays, databases and data mining tools, collections of knockout and insertion mutants, clones in various expression vectors, and even protein microarrays. The community of researchers studying *Xanthomonas* is growing and many have in the last decade embraced powerful genomic-technologies to better understand how this important group of bacteria has adapted to exploit an extraordinary diversity of plants. Utilizing such techniques, *Xanthomonas* researchers are collecting large amounts of diverse genomic data with ever-increasing speed. The exponential growth of information generated by this type of research has many challenges associated with it. To maintain, update, and make accessible these large sets of data, it has become critical for research communities to put forth a concerted effort rather than rely on the often uncoordinated activities of a few groups.

Discussion participants set out to identify the most important issues now facing the *Xanthomonas* community in this regard. Points raised centered on two major issues: (1) management, sharing, and access to genomic data and materials, and (2) communication within the community and to the wider scientific community. Examples were cited of species- or subject-specific web based platforms for facilitating dialog among research groups, in particular allowing people with differing expertise to bridge gaps between genomics, bioinformatics, and experimental lab research.

All agreed that a web based central portal through which researchers in the community could exchange information (protocols, data, software, announcements of meetings and funding opportunities, advice(!), etc.) and that would facilitate sharing of materials, would be ideal. It was suggested that this portal could also be set up to provide easy access to databases for genomic sequence and expression data and associated tools for mining and graphical representation. The portal could also provide or link to related data and resources such as type strain collections, mutant libraries, cloning and expression vectors, as well as emerging technologies and software as they are developed.

Managing and maintaining such a resource is a major challenge for an international community of researchers whose available funding and personnel change over time, and for which coordinated, long-term, international funding programs are not in place. Contributing to such a resource would be for many a service activity competing for time with research and teaching responsibilities. And even with the good will and cooperation of numerous groups, clearly one or a few groups would need to step forward to take the lead. In spite of these challenges, because a community platform has such potential to accelerate advances in *Xanthomonas* biology, and because there is strong community support for the concept, the meeting participants expressed optimism that significant progress toward this goal will be made in the near future. Indeed, [www.xanthomonas.org](http://www.xanthomonas.org) has recently been established by Dr. Ralf Koebnik, housing links to genome data and tools, contact information for researchers by area of expertise, gene ontology and taxonomy data, and a nascent forum for sharing news.

## Conclusion

Though detailed guidelines remain to be established, the need for a rational, unified approach to nomenclature for *Xanthomonas* type III effector proteins was recognized, and a general framework was established. Key to the framework was consensus that gene names well-established in the literature should be retained, and that source should somehow be linked to names of specific genes. Sequencing priorities varied among participants, but a majority supported targeting epiphytic or non-pathogenic *Xanthomonas* populations as well as representative stains for each of the currently non-sequenced species. At the same time, the group recognized the importance of sequencing deeply within a few taxonomic groups that are well-studied. Momentum has begun to build for a web-based community platform for sharing information, data, and materials. A new website was announced that has many useful features. Further development will benefit from continued community engagement and contribution, and support for this type of activity from public research funding agencies. The authors look forward with excitement to the progress and new challenges to be discussed at the next *Xanthomonas* Genomics Conference, in 2012.

## References

- Almeida, N.F., Yan, S., Lindeberg, M., Studholme, D.J., Schneider, D.J., Condon, B., Liu, H., Viana, C.J., Warren, A., Evans, C., Kemen, E., Maclean, D., Angot, A., Martin, G.B., Jones, J.D., Collmer, A., Setubal, J.C., and Vinatzer, B.A. 2009. A draft genome sequence of *Pseudomonas syringae* pv. tomato T1 reveals a type III effector repertoire significantly divergent from that of *Pseudomonas syringae* pv. tomato DC3000. *Mol Plant Microbe Interact* 22:52-62.
- Aury, J.M., Cruaud, C., Barbe, V., Rogier, O., Mangenot, S., Samson, G., Poulain, J., Anthouard, V., Scarpelli, C., Artiguenave, F., and Wincker, P. 2008. High quality draft sequences for prokaryotic genomes using a mix of new sequencing technologies. *BMC Genomics* 9:603.
- Guo, M., Tian, F., Wamboldt, Y., and Alfano, J.R. 2009. The majority of the type III effector inventory of *Pseudomonas syringae* pv. tomato DC3000 can suppress plant immunity. *Mol Plant Microbe Interact* 22:1069-1080.
- MacLean, D., Jones, J.D., and Studholme, D.J. 2009. Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol* 7:287-296.
- Pallen, M.J., and Wren, B.W. 2007. Bacterial pathogenomics. *Nature* 449:835-842.
- Reinhardt, J.A., Baltrus, D.A., Nishimura, M.T., Jeck, W.R., Jones, C.D., and Dangl, J.L. 2009. De novo assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae* pv. *oryzae*. *Genome Res* 19:294-305.
- Salzberg, S.L., Sommer, D.D., Schatz, M.C., Phillippy, A.M., Rabinowicz, P.D., Tsuge, S., Furutani, A., Ochiai, H., Delcher, A.L., Kelley, D., Madupu, R., Puiu, D., Radune, D., Shumway, M., Trapnell, C., Aparna, G., Jha, G., Pandey, A., Patil, P.B., Ishihara, H., Meyer, D.F., Szurek, B., Verdier, V., Koebnik, R., Dow, J.M., Ryan, R.P., Hirata, H., Tsuyumu, S., Won Lee, S., Seo,

Y.S., Sriariyanum, M., Ronald, P.C., Sonti, R.V., Van Sluys, M.A., Leach, J.E., White, F.F., and Bogdanove, A.J. 2008. Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. *BMC Genomics* 9:204.

Matthieu Arlat

Anke Becker

Adam Bogdanove\*

Ralf Koebnik

Jan Leach

Mary Beth Mudgett

Robert Ryan

Boris Szurek

Frank White

\* for correspondence, [ajbog@iastate.edu](mailto:ajbog@iastate.edu)