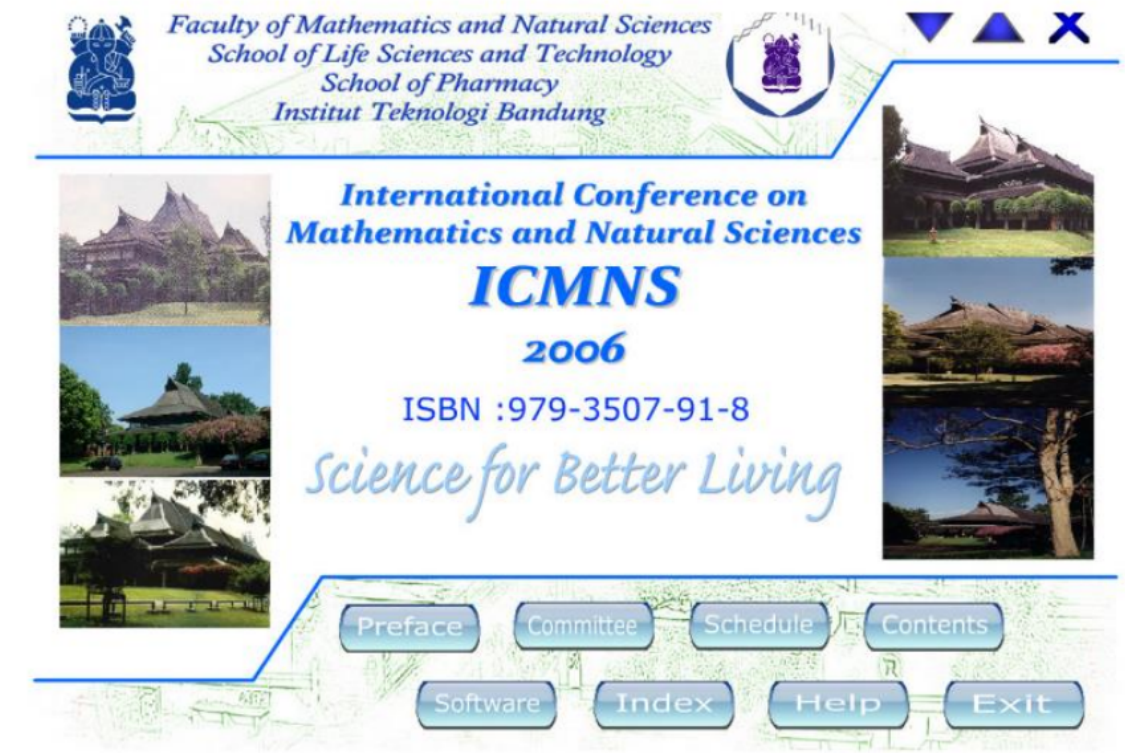
COMPARISON OF THE GENOMES OF FOUR Xanthomonas PATHOGEN STRAINS

By Marieska Verawaty

WORD COUNT 2515
TIME SUBMITTED 30-APR-2023 06:57PM
PAPER ID 99232649


Faculty of Mathematics and Natural Sciences
School of Life Sciences and Technology
School of Pharmacy
Institut Teknologi Bandung
International Conference on
Mathematics and Natural Sciences
ICMNS
2006
ISBN :979-3507-91-8
Science for Better Living
Preface
Committee
Schedule
Contents
Software
Index
Help
Exit

Preface

ICMNS was held in ITB Campus, Bandung, 29 – 30 November 2006, under the joint auspices of Faculty of Mathematics and Natural Sciences, School of Pharmacy, and School of Life Sciences and Technology, Institut Teknologi Bandung.

We have been, almost, overwhelmed by the level of interest and participation in the conference. Overall, 13 invited lectures were delivered and 526 papers were presented (212 oral presentations and 314 poster presentations). The proceedings published in a CD-ROM form.

Space is unfortunately too short to thank individually all those people whose dedicated efforts, often over a period of many months, were responsible for the success of the conference - especially the members of the Steering and Organizing Committees.

Ismunandar

Conference Chairman

Back

COMPARISON OF THE GENOMES OF FOUR *Xanthomonas* PATHOGEN STRAINS

Marieska Verawaty [1], Adam J. Bogdanove [2] and Volker Brendel [3]

1. **Biology Department, Sriwijaya University, Ogan Ilir, South Sumatera, Indonesia.**
2. **Plant Pathology Department, Iowa State University, Ames, IA 50011, USA**
3. **Bioinformatics, Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011, USA**

ABSTRACT

Genome comparison of Xanthomonas oryzae pv oryzicola (Xoc) with Xanthomonas campestris pv campestris (Xcc), Xanthomonas axonopodis pv citri (Xac) and Xanthomonas oryzae pv oryzae (Xoo) had been analyzed in Bioinformatics and Plant Pathology Department, Iowa State University, USA. Xac and Xoc both are non vascular host type infection; Xac infected dicots host but Xoc infected monocots host. On the other hand Xcc and Xoo both are vascular host type infection; Xcc infected dycot host but Xoo infected monocot host. The aim of this research was to find the unique genes both in vascular and non vascular host type infection and the unique genes both in dicots and monocots host type classification. On this research Open Reading Frames (ORFs) were identified using dnatopro. Sequence similarity searches were done with stand alone Basic Local Alignment Search Tool (BLAST) and NET-BLAST. Multi-query Sequence Box (MuSeqBox) was using to parse BLAST output and store attributes of BLAST hits in tabular form. It was found that there were 3,085 qualifying ORFs in the Xoc genome sequence with minimal ORF length set to 300 amino acids. From this research we found that there are 318 candidate non-vascular specific proteins. There are 361 candidates of vascular specific proteins. There are 572 candidates of dicots specific proteins. There are 1097 candidate monocots specific proteins. And also we found that there are 58 proteins potentially really unique in the Non-Vascular species and 134 proteins potentially really unique in the vascular species.

Keywords : *Genomic comparison, Xanthomonas sp., Vascular, non vascular, dicots and monocots unique genes, BLAST.*

I. Introduction

Xanthomonas is a large genus of Gram-negative, yellow-pigmented, plant associated bacteria that show a high degree of host plant specificity. Pathogenic members of the genus together cause diseases on over 392 plant hosts (Hayward, 1993) including both dicots (pepper to citrus) and monocots (rice to wheat). Many exhibit tissue-specific pathogenicity. By comparing genomes of different Xanthomonas species and pathovars, we could begin to understand fundamental aspects of bacterial plant pathogenicity, including what genes are important for host-species specificity and what genes are required for invasion of different tissues, or for pathogenesis generally, irrespective of host.

Xoc is an important non-vascular pathogen of the crop and model monocot rice. It is closely related to Xoo, the vascular pathogen of the same host that has already been sequenced (Cook, et al 1952; Bretschneider, et al 1989). Vascular and non-vascular infections are distinct modes of pathogenesis together representative of the majority of bacterial plant diseases. Comparisons at different levels would uncover genes (or gene sets) that are found only in the vascular pathogens, only in the non-vascular colonizer, or only in the pathogens of a one type of host or the other. Functional characterization of these genes will benefit from the unique opportunity that these specific pathogens afford to work in model hosts. The Xanthomonas genome sequences would be a valuable resource for a large community of researchers, and the availability of the sequences would contribute substantially to our understanding of bacterial interactions with plants and to the sustainability and safety of agriculture.

DATA and PROGRAMS

DATA: The genome sequences of three Xanthomonas strains and associated annotations were downloaded from NCBI, http://www.ncbi.nlm.nih.gov/genome/ with accessions numbers: NC_006834 (Xoo), NC_003919 (Xac) and NC_003902 (Xcc). The Institute of Genomic Research (TIGR) provided the Xoc genome in draft form.

PROGRAMS: Open Reading Frames (ORFs) were identified using dnatopro (V. Brendel, unpublished). Sequence similarity searches were done with stand alone Basic Local Alignment Search Tool (BLAST) and NET-BLAST. Mul query Sequence Box (MuSeqBox) was using to parse BLAST output and store attributes of BLAST hits in tabular form (Xing & Brendel, 2001).

I. First assembly of the Xoc genome sequence. By using the dnatopro program with minimal ORF length set to 300 amino acids, we found that there are 3,085 qualifying ORFs in the Xoc genome sequence.

The number of ORFs of Xoc that correspond to annotated proteins of Xac, Xcc and Xoo was found by Searching steps against compiled database using BLASTP (E-value:1e-20), it was found about 1,137 (36,86%) Xoc ORFs with no close similarities with the annotated compiled proteins.

The comparison of these ORFs against all known proteins (NCBI non redundant data base) using NET-BLAST, it was found 971 (31,48%) unique ORFs of Xoc that didn't hit anything and 166 (14,6%) protein hits with non redundant data base (nr). The unique ORFs in Xoc that hit with nr NCBI that match in the Xac, Xcc and Xoo nucleotide sequence were shown in Table 1.

Table 1. TBLASTN Results of Xoc ORFs against the Xac, Xcc, and Xoo genome sequences.

NO	TBLASTN	Total Query	Total Hits	Total No Hits
1.	orfXoc-nr-vs-XacFAS	166	101	65
2.	orfXoc-nr-vs-XccFAS	166	99	67
3.	orfXoc-nr-vs-XooFAS	166	142	24

From the TBLASTN results, we conclude that the 971 Xoc's unique ORFs are most likely random ORFs, 142 ORFs may be diverged proteins (giving low BLASTP scores) or unannoted (unlikely) proteins in compiled database, and the other 24-maybe real genes.

III. Comparison of Xoc new assembly genome sequence with Xcc, Xac and Xoo genomes.

The following results were based on the second assembly of the Xoc genome provided by The Institute of Genomic Research (TIGR) as Xoc 4.Bigscaff.fasta (its length was 4,813,407 bp). By using TBLASTN at level le-20, it was found that there were 1,905 proteins in Xac, 816 proteins in Xcc and 317 proteins found in Xoo that are not found in Xoc. These are the proteins seems to be unique to Xac, Xcc and Xoo relative to Xoc.

Common genes in Xac, Xcc and Xoo that are not found in Xoc was found by using TBLASTN, the results were shown in Table 2.

Table 2. TBLASTN results of common gene in Xac,Xcc and Xoo against Xoc.fas

TBLASTN	3-way common	4-way common	No Hits
cXacXooXcc-vs-Xoc.fas	3087	3017	70
cXccXacXoo-vs-Xoc.fas	3085	2995	70
cXooXccXac-vs-Xoc.fas	3151	2986	66

c refer to concatenated (3-way common), then its results was searched against Xoc.fas using TBLASTN (4-way common).

From the TBLASTN results in the Table 4, we concluded that there are about 69% genes present in the four genomes of *Xanthomonas* and about 70 genes/proteins (2,33%) common to Xac/Xoo/Xcc that may not be present in Xoc. The 70 genes are listed bellow:

acetyltransferas
microcyctin dependent protehypothetical protein
L-fucose dehydse sugar-phosphate isomerase
lytic enzyme
CDP-alcohol-phosphatidil transferase
anti-sigma Factor antagonist
Icf G protein
arabinogalactan endo 1,4 beta galactosidase
truncated xylanase
gluthathione S-transferase
Outer Membrane component of multidrug efflux pump
general stress protein
glycosyltransferase
Site specific DNA methyltransferase
Phage related capsid packaging proteinPhage related terminase
Phage related capsid protein
Phage related major capsid protease
Phage related capsid compprotein
Phage related tail protein
Phage related lytic enzyme
Phage related bass plate assembly protein
methylated DNA protein
cystein S methyltransferase related protein
type I restriction modification system DNA methylase
OMP W
OMP
truncated cellulose
Oxidoreductase
RTS beta protein
3-dehydroquinqte dehydratase
bleomycine resistance protein
membrane protein
inner membrane protein
Glutatione transferase
Cellulase S
arabinogalactan
two component sensor protein
transcriptional regulator
cardiolipin synthase
ton B dependent receptor
Endopolygalacturonase

III.1. COMPARISON OF STRAIN SPECIFIC (VASCULAR-vs-NON VASCULAR AND DICOTS vs MONOCOTS)

IN THE FOUR *Xanthomonas* GENOME SEQUENCES.

Potentially unique genes were searched to understand vascular and non-vascular bacterial diseases of both dicots and monocots. It can be drawn that there are four type of species specific groups on this research; (1) Non Vascular (NV): Xac and Xoc, (2) Vascular (V): Xcc and Xoo, (3) Dicot (D): Xac and Xcc, and (4) Monocot (M): Xoc and Xoo.

Table. 3. General features of Xac, Xcc, Xoo and Xoc genomes

Strains	Genome size (MB)	Tot CDS	GC content (%)
Xac	5.17	4312	64.7
Xcc	5.07	4181	65.0
Xoo	4.94	4637	63.7
Xoc	4.81	4287	64.1

To search the potentially unique genes to the NV, V, D and M group. It was conducted the following BLASTP searches: 1. NV (Xac-Xoc) against V (Xcc-Xoo) to find genes potentially unique to Non-Vascular. 2. V (Xcc-Xoo) against NV (Xac-Xoc) to find genes potentially unique to Vascular. 3. D (Xac-Xcc) against M (Xoo-Xoc) to find genes potentially unique to Dicot. 4. M (Xoo-Xoc) against D (Xac-Xcc) to find genes potentially unique to Monocot.
The BLASTP results are shown in Table 4.

Table 4. BLASTP results of comparison among the four groups species specific in the *Xanthomonas* genome sequences.

Species Specific	**Dicot**	**Monocot**	**Total genes**
Non Vasc	Xac	Xoc	3401 com = ***318 uNV*** + 3083
Vascular	Xcc	Xoo	3464 common = ***361 uV*** + 3102
Total Genes	3581 com = ***572 uD*** + 3009	4139 common = ***1097 uM*** + 3042	

There are 318 proteins are the candidate non-vascular specific proteins, 361 candidates of vascular specific proteins, 572 proteins are candidates of dicot specific proteins, and 1097 proteins are the candidate monocot specific proteins.

The proteins in uNV that have no TBLASTN hit to the Xcc and Xoo genomes are the potential proteins to be really unique Non-Vascular (ruNV). There are 361 proteins unique in V (uV). The proteins in uV that have no TBLASTN hit to the Xac and Xoc genomes are the potential protein to be really unique Vascular (labeled ruV). To explore this further, we conducted the following TBLASTN searches: (1) uNV against Xcc.fas. (2) uNV against Xoo.fas. (3) uV against Xac.fas. (4) uV against Xoc.fas. This allows us to break up the uNV set into four parts: protein present in both Xcc and Xoo or missing in either Xcc or Xoo or in both. The results are shown in Table 5 and 6.

Table 5.TBLASTN of uNV against Xcc.fas and Xoo.fas

Label	**Total genes**	**Xcc.fas**	**Xoo.fas**
nruNV	57	y	y
	79	y	n
	124	n	y
ruNV	58	n	n
Total	318		

Table 6. TBLASTN of uV against Xac.fas and Xoc.fas

Label	**Total genes**	**Xac.fas**	**Xoc.fas**
nruV	16	y	y
	139	y	n
	72	n	y
ruV	134	n	n
Total	361		

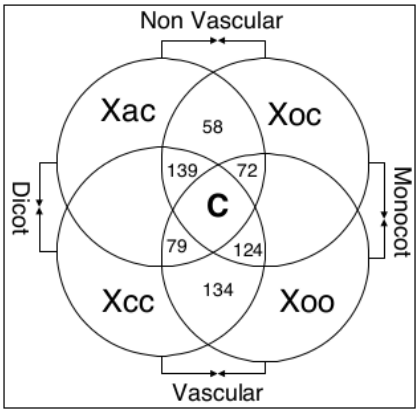


Figure 1. Venn diagram of distribution of the potentially really unique genes in vascular, non vascular, dicot and monocot *Xanthomonas* host's species specificity. (C refered to common genes: it was about 3050 genes totally).

The Venn diagram in Figure.1 shown that there are about 3050 genes present in the four Xanthomonas genome sequences, 58 (1,93%) genes potentially really unique NV and about 134 (3,03%) genes really unique Vascular. The list of the 58 potential proteins can be seen below:

50 S ribosomal protein L34	OMP
ABC transporter ATP bind prot	Oxidoreductase
Acylphosphatase	Peptidase
Fimbrial biogenesis protein	Phytoene desaturase
Hemagglutinin	Plasmid mobilization protein
Histone H1	Transcriptional Regulator
Hpa1 protein	Transglycosylase associ pro
Integral membrane protein	UDP-N-acetyl-D-mannosamine transferase
NAD dependent epimeras	

The 134 proteins potentially really unique in the vascular species are belong to: acetyltransferase, alcohol dehydrogenase, aminopeptidase N, AttT protein, isomerase, flagelar protein, histidine kinase, histone H1, *Hpa1, HrpE, HsdR*, ISD1 transposase, Isopenicillin N epimerase, ISxcd1 transposase, leucin rich protein, methyltransferase, nisin resistance protein, nodulation related protein, oxidoreductase, pectinesterase, phage-related baseplate assembly protein, phage related tail protein, pilin, polymerase V subunit, putative DNA helycase, RNA-pol sigma-70 factor, transaminase, transcriptional reg araC family, truncated IS1477 transposase, WxcH protein, and *XmnI* methyltransferase.

This comparative analyses between the four *Xanthomonas* strains help us to identify a set of strain-specific genes, some of which are probably responsible for distinct pathogenecity and host specificity profiles of these organisms. On this genomic approach, the unique genes shown some involvements in kinds of pathogenecity such as: cell-cell interactions (fimbrial, hemagglutinin-like proteins), plant cell wall degradation (cellulases, xylanases, pectinases, proteases), Iron homeostasis (genes encoding receptors for several siderophores and protein involved in iron siderophore transport across the membrane: *tonB*, genes involved in type III protein secretion system, and in plant pathogens (*hrp*: for hysensitive reaction and pathogenecity). *Hrp* gene suggesting that their products are delivered via the type III secretion system; *Hrp* gene was found in ruV and dicots unique genes (lambais, et al. 2000).

REFERENCES

Bretschneider, K.E., Gonella, M.P. & Robeson, D.J. 1989. A comparative light and electron microscopical study of compatible interactions between *Xanthomonas campestris* pv campestris and cabbage (Brassica oleracea). *Physiol Mol Plant Pathol* 34, 285-297.

Cook, A., Walker, J. & Larson, R. 1952. Studies on the disease cycle of black rot of crucifers. *Phytopathology* 42, 162-167.

Da Silva, A. C., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Sluys., M.A., Almeida, N.F., Alves, L.M., Do Amaral, A.M., Bertolini, M. C., Camargo, L. E., Cammarote, G., Cannavan, F., Cardozo,., Chambergo, F., Ciapina, L.P., Cicarelli, R.M., Coutinho, L.L., Cursino-Santos, J.R., El-Dorry, H., Faria, J. B., Ferreira, A.J., Ferreira, R. C., Ferro, M. I., Formighieri, E.F., Franco, M. C., Greggio, C. C., Gruber, A., Katsuyama, A.M., Kishi, L.T., leite, R.P., Lemos, E. G., Lemos M.V., Localli, E.C., Menck, C.F., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T., Okura, V.K., Oliveira, M.C., Oliveira, V.R., H.A., Rossi, N.M., Novo, M.T., Okura, V.K., Sena, J.A., Silva, C., De Souza, R.F., Spinola, L.A., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I., Trindade Dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C.. & Kitajima, J.P. 2002. Comparison of the genomes of two Xanthomonas pathogens with differing host specificities. *Nature* 417, 459-63.

Donga, Q. & Brendel V. 2004. Computational identification of relates proteins BLAST, PSI-BLAST, and Other Tools. *The Proteomics Protocols Handbook*. Edited by: J.M. Walker @ Humana Press Inc., Totowa, NJ.

Hayward, A. 1993. The hosts of Xanthomonas *in Xanthomonas*, eds. Swings, J. & Civerelo, E.L. (Chapman and Hall, London), pp. 1-18.

Lambais, M.R., Goldman, M.H.S. Camargo, L.E.A., and Goldman, G.H. 2000. A genomic approach to the understanding of *Xilella fastidiosa* pathogenecity. *Curr Opin in Microb,* 3:459-462.

MuSeqBox,http://bioinformatics.iastate.edu/bioinformatics2go/mb/MuSeqBox.html.

http://www.ncbi.nlm.nih.gov/genome/

Xing, L. & Brendel, V. 2001. Multi-query sequence BLAST output examination with MuSeqBox. *Bioinformatics Applications Note,* 17: 744-45.

ACKNOWLEDGEMENTS.

This work was supported by Bergdhal Endowment for Prof. Volker Brendel, and Plant Pathology Research Project, USDA grants for Prof. Adam J. Bogdanove during the short training course in Iowa State University, I wish to thank to Brendel's group for the assistances, and for The Institute of Genomic Research (TIGR).

COMPARISON OF THE GENOMES OF FOUR Xanthomonas PATHOGEN STRAINS

ORIGINALITY REPORT

5%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet	38 words — 2%
2	www.damienmeyer.fr Internet	28 words — 1%
3	Dyann F. Wirth. "The parasite genome: Biological revelations", Nature, 10/03/2002 Crossref	24 words — 1%
4	mafiadoc.com Internet	12 words — 1%
5	Lu, Hong. "Application of comparative genomics for detection of genomic features and transcriptional regulatory elements", Proquest, 20111109 ProQuest	11 words — 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 1%
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 1 WORDS