GENOME SEQUENCE OF A STRAIN OF *XYLELLA FASTIDIOSA* ASSOCIATED WITH PIERCE'S DISEASE IN CALIFORNIA

Project Leader:
Edwin L. Civerolo
USDA, ARS
Davis, CA 95615

Cooperators:

Marie-Anne van Sluys	*Mariana C. Oliveira*	*Joao Paulo Kitajima*
Instituto de Biociencias	*Instituto de Biociencias*	*Laboratorio de Bioinformatica*
Sao Paulo, Brazil	*Sao Paulo, Brazil*	*Campinas, Brazil*

FAPESP, ONSA
Agronomical and Environmental Genomics
Sao Paulo, Brazil

INTRODUCTION

Several economically important diseases of agronomic and horticultural crops, as well as landscape and forest trees, are caused by different strains or pathogenic variants of *Xylella fastidiosa* (*Xf*). In California, these include (but are not necessarily limited to) alfalfa dwarf, almond leaf scorch, oleander leaf scorch, and Pierce's disease (PD). Other *Xf*-caused diseases that are potential threats to California agriculture are citrus variegated chlorosis (CVC), phony peach, and plum leaf scald. Recently, the complete genome sequence of *Xf* strain 9a5c, which causes CVC in Brazil, was determined (3). Such information is useful for increased understanding of *Xf*-host interactions in order to develop new disease management strategies. However, the nature of, and mechanism(s) involved in, differing host ranges of *Xf* strains are not completely understood. Therefore, comparative information about the genome structure, specifically the complete genome sequence, of another *Xf* strain (besides *Xf*-CVC strain 9a5c) could contribute to elucidation of *Xf*-host and *Xf*-insect vector interactions. Accordingly, the complete genome sequence of a strain of *Xf* associated with PD in California was determined through a cooperative project between the USDA-ARS (including AVF and CDFA) and FAPESP.

OBJECTIVE

Determine the complete genome sequence of a *Xylella fastidiosa* strain associated with Pierce's disease in California.

RESULTS AND CONCLUSIONS

Genome assembly of a strain of *Xf* associated with PD in California (*Xf*-PD [Temecula1]) was achieved through shotgun and cosmid sequencing by the AEG network from the ONSA-FAPESP program. The draft generated so far has 2,507,178 bp, which corresponds to 93.6 % of the previously sequenced *Xf* strain associated with citrus variegated chlorosis (*Xf*-CVC). In addition, only a miniplasmid consisting of 1.345 bp was found in the *Xf*-PD (Temecula1) genome. This plasmid is similar to the previously described miniplasmid in the genomes of other *Xf* strains.

Xf-PD v1.0 draft genome assembly was obtained after using "Scaffold" program (2) on our database consisting of 191 contigs generated from shotgun reads and cosmid ends. From the scaffolded genome, 14 cosmids and 2 plasmids were selected for further sequencing in order to close the genome. The overall base quality achieved for this genome is phred quality above 50 for 99.7 % of the nucleotide bases. Low quality regions are being screened to determine regions to have base quality improved.

The *Xf*-PD genome harbors large rearrangements compared to the *Xf*-CVC genome. These rearrangements include some translocation events as well as insertion/deletion (INDEL) events. INDEL events can either be represented by single missing or inserted gene or sets of genes, most being related to phage sequences. These structural differences were first identified by cross match analysis and the genes involved in these rearrangements are now being identified by the annotation of the genome.

Open reading frames (ORFs) were identified in the genome using the "GeneMark" program (1) and subsequently analyzed by a group of 40 annotators. Annotation is being carried in a first set of 1,935 identified ORFs. From these, 6.0% are exclusively ORFs from the *Xf*-PD genome while the *Xf*-CVC genome has 25% exclusive ORFs. Both the *Xf*-PD and *Xf*-CVC genomes have similar distributions of mobile genetic elements, which include phage, transposon and plasmid-related sequences.

There are ORFs in the *Xf*-PD genome that are present in more copies than in the *Xf*-CVC genome. Also, there are missing genes in the *Xf*-CVC genome that are missing in the *Xf*-PD genome. These are being carefully analyzed and the final description is dependent upon completion of the annotation process. Interestingly, there is a region in the *Xf*-CVC genome, which is missing in the *Xf*-PD genome. This region encompasses 76 genes corresponding to at least a deletion of 70,000 bp. Also, no group II intron was detected in the *Xf*-PD genome. On the other hand, there is at least one new phage insertion in the *Xf*-PD genome not described in *Xf*-CVC genome. These comparative analyses will help target specific genes in either strain to be studied as well as common genes that enable these bacteria to live in the plant xylem and in the insect foregut.

REFERENCES

Borodovsky, M. and J.D. McIninch. 1993. GeneMark: Parallel gene recognition for both DNA strands. Chemistry 17: 123-133.

Setubal, J.C. and R. F. Werneck. 2001. A program for building contig scaffolds in double-barreled shotgun genome sequencing. UNICAMP Computing Institute Technical Reports #IC-01-05, 20 pp.

Simpson, et al. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. Nature 406:151-159.