Plant Omics Journal POJ

POJ 5(5):494-502 (2012) ISSN:1836-3644

In silico characterization and molecular modeling of GntR family regulators in *Xanthomonas axonopodis* pv. *citri*: Implications for primary metabolism or virulence

Liping Liu†, Jinwang Qu†, Na Li, Dazhi Li, Guiyou Long and Ziniu Deng*

Hunan Provincial Key Laboratory of Crop Germplasm Innovation and Utilization, Hunan Agricultural University, Changsha 410128, China
National Center for Citrus Improvement, Hunan Agricultural University, Changsha 410128, China

†Authors contributed equally to this work.

***Corresponding author: deng7009@163.com**

Abstract

Xanthomonas axonopodis pv. *citri* (*Xac*) causes citrus canker, a serious threat to the citrus industry worldwide. The GntR family is a group of helix-turn-helix bacterial transcriptional regulators, and may be involved in bacterial virulence or primary metabolism. In this study, genome-wide sequence analysis of *Xac* strain 306 revealed that two genes encode the FadR-like regulators. Moreover, one gene encodes the YtrA-like regulator, one gene MocR-like regulator, and three the HutC-like regulators. Furthermore, an attempt was done to predict the 3D models of these regulators using I-TASSER server. Backbone conformation of the modeled structure by PROCHECK revealed that more than 95% of the residues fall in the allowed regions. ProQ results confirmed the high quality of modeled structure and verified by "Verify 3D" profiles at satisfying level. Considering the auto-regulatory characteristic of GntR-like genes, the potential binding sites were identified by analyzing the primary amino acid sequences. On the basis of reciprocal BLAST search, GntR orthologs in other members of the genus *Xanthomonas* were also identified. Our analysis provides valuable clues for initiation of experimental characterization of these regulators and enhances our understanding about their roles in citrus-*Xac* interactions.

Keywords: citrus canker; *Xanthomonas axonopodis* pv. *citri*; GntR regulator; virulence; metabolism; homology modeling.
Abbreviations: 3D_three-dimensional; E-b/O_effector binding and/or oligomerization; I-TASSER_iterative threading assembly refinement server; PLP_pyridoxal 5'-phosphate; PMDB_Protein Model DataBase; PMP_pyridoxamine 5'-phosphate; *Xac_Xanthomonas axonopodis* pv. *citri*.

Introduction

The gram-negative, rod-shaped γ-proteobacterium *Xanthomonas axonopodis* pv. *citri* (*Xac*) is the causal agent of citrus canker, a major threat to citrus production worldwide. *Xac* can infect almost all citrus plants, including oranges, sour oranges, grapefruits, tangerines, lemons, and limes (Das, 2003). This organism has been widely used as a model to study the molecular mechanisms of pathogen-plant interactions (Graham et al., 2004; Ryan et al., 2011; Liu et al., 2012). *Xac* can cause eruptive, pustule-like lesions with water-soaked margin often surrounded by a chlorotic halo on citrus leaves, fruits and stems. For *Xac*, overcoming the multilayer plant defense fortification and acquiring nutrients from citrus plant is the prerequisite for successful infection and proliferation during the pathogenic process. Over the past decade, lots of genes that essential for the virulence of *Xac* have been identified (Lu et al., 2008; Laia et al., 2009; Malamud et al., 2011; Li and Wang, 2012), including gene *pthA*, which is required for pathogenicity of *Xac* (Swarup et al., 1991; Al-Saadi et al., 2007). The GntR family is an abundant group of helix-turn-helix bacterial transcriptional regulators (>32500 members in the Pfam database; Mar 2012). It was named after the first identified *Bacillus subtilis* repressor of the gluconate operon (Haydon and Guest, 1991). Regulators of this family comprise an N-terminal winged helix-turn-helix domain, followed by a C-terminal effector-binding and/or oligomerization domains (E-b/O) involved in ligands binding. However, the C-terminal domain is quite heterogeneous and diverse. As a consequence of this heterogeneity, the GntR regulators have been further categorized into six specific subfamilies, namely, the FadR, HutC, MocR, YtrA, AraR, and PlmA subfamilies (Lee et al., 2003; Rigali et al., 2004). The members of subfamilies own conserved subfamily-specific secondary structural elements and can interact with several molecules (Rigali et al., 2002). It was reported that many important GntR family members regulate the most varied biological processes (Rigali et al., 2002; Hoskisson and Rigali, 2009). A large repertoire of characterized GntR regulators is associated with various metabolic pathways (Yoshida et al., 2000; Frunzke et al., 2008; Hyeon et al., 2012). However, thus far, only three pieces of evidence indicate that GntR members can regulate virulence in pathogenic bacteria recently. First, systematic targeted mutagenesis of *Brucella melitensis* 16M revealed a major role for GntR regulators in the control of virulence (Haine et al., 2005); second, Mce1R, a homolog of the FadR subfamily, is required for organized granuloma formation, which is both protective to the host and necessary for the persistence of *Mycobacterium tuberculosis* (Casali et al., 2006); and third, systematic mutagenesis of all predicted *gntR* genes in *X. campestris* pv. *campestris* revealed a GntR family transcriptional regulator controlling hypersensitive response

Table 1. Basic information of seven classified *Xac* GntR-like regulators.

Regulator name	Subfamily	Protein size (amino acids)	UniProt ID
XAC0568	FadR	239	Q8PPW5
XAC0711	HutC	255	Q8PPH5
XAC0737	MocR	466	Q8PPE9
XAC0877	FadR	224	Q8PP11
XAC1548	YtrA	120	Q8PM84
XAC1640	HutC	234	Q8PLZ5
XAC3532	HutC	242	Q8PGT4

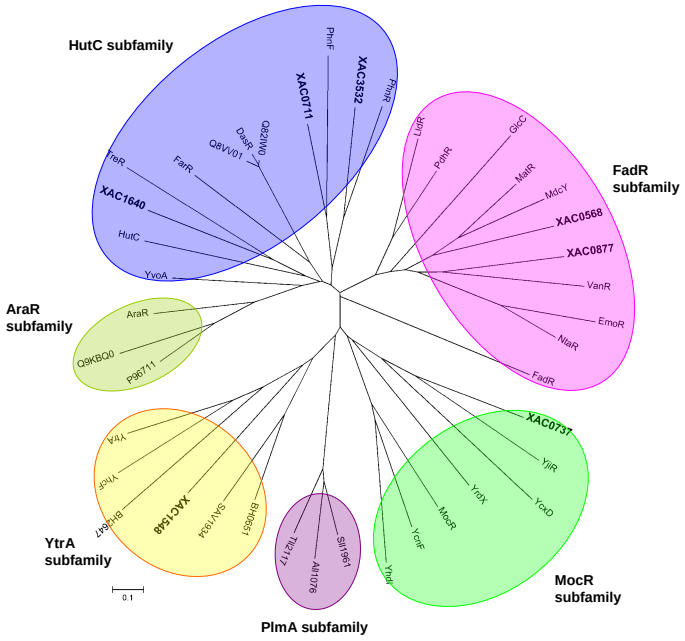

Fig 1. Classification of *Xac* putative GntR-like regulators. An unrooted phylogenetic tree was generated by neighbor-joining (NJ) method with MEGA5 based on the predicted amino acid sequences of the putative GntR-like transcriptional regulators in *Xac* (highlighted in ***bold***) and their homologues in other species. The branch lengths are proportional to divergence, with the scale of '0.1' representing 10% change. Details of GntR regulators used as representatives from all subfamilies are given in Supplemental Table S1.

and virulence (An et al., 2011). The whole genome of *Xac* strain 306 has been completely sequenced and deposited in the National Center for Biotechnology Information (NCBI) database, which contains a number of putative GntR-like regulators (da Silva et al., 2002). This information suggests that the GntR family of transcriptional regulators may play an important role in symbiosis between *Xac* and citrus. Recently, several genome-level studies on the distribution of GntR regulators have been published, and most of them focused on *Mycobacterium* (Vindal et al., 2007a; Vindal et al., 2007b; Ji and Xie, 2011), while there was no report on genome-wide studies specifically addressed the characteristics of the three-dimensional (3D) modeling and binding site residues of GntR factors in *Xanthomonas*.

In the present study, seven putative GntR regulators from *Xac* were classified into four specific subfamilies. The 3D modeling of these regulators was constructed based on the structure characters and validated with standard parameters (PROCHECK, ProQ and Verify 3D). Furthermore, suitable orthologs of the *Xac* GntRs were also identified using reciprocal BLAST searches in *Xanthomonas* spp.. This study could be useful in further functional characterization of this important group of bacterial transcriptional regulators.

Results and discussion

Classification of the putative Xac *GntRs into subfamilies*

To classify putative GntR regulators into subfamilies, an unrooted tree of putative *Xac* GntRs was constructed with other known and classified representatives of all subfamilies (Table 1) (Rigali et al., 2002). As shown in Fig. 1, all seven putative members of the *Xac* GntR family have been classified into four subfamilies FadR, YtrA, MocR, and HutC, but none of them were clustered with other GntR subfamilies such as AraR and PlmA.

FadR-like regulators of Xac

Of all seven putative *Xac* GntRs, two proteins (XAC0568 and XAC0877) were classified as the FadR-like regulators. This subfamily possesses all α-helical C-terminal domains with an average length of about 160 amino acids. According to previous study, these subfamily members are further classified into two groups, FadR and VanR. Regulators with six α-helices C-terminal domain were classified into VanR group, while regulators with seven α-helices were classified into FadR group (Rigali et al., 2002). XAC0568 and

Table 2. Ramachandran plot statistics for the 3D model of *Xac* GntR-like regulators, calculated using PROCHECK and ProQ servers.

Modeled regulator	Template PDB (chain)[#]	PMDB ID	PROCHECK parameter (%)			G factor	LGscore[*]
			Favored	Allowed	Outlier		
XAC0568	3C7J (A)	PM0077807	91.6	3.8	4.6	0.01	3.077
XAC0711	2WV0 (D)	PM0077808	89.3	7.9	2.8	-0.25	2.748
XAC0737	1WST (A)	PM0077809	92.2	5.6	2.2	-0.22	3.148
XAC0877	2HS5 (A)	PM0077810	94.6	3.2	2.2	0.05	2.438
XAC1548	3BY6 (E)	PM0077811	94.1	5.1	0.8	0.02	1.865
XAC1640	2WV0 (D)	PM0077812	94.9	3.4	1.7	-0.18	1.546
XAC3532	2WV0 (D)	PM0077813	93.8	4.6	1.6	-0.16	1.580

[#] Top one template used by I-TASSER. [*] Different ranges of quality: LGscore>1.5 fairly good model, LGscore>2.5 very good model (Cristobal et al., 2001).

XAC0877 were predicted having six α-helices C-terminal domains and both of them showed distinguishable predicted secondary structural features specific to FadR subfamily (Fig. 2a) (Rigali et al., 2002). So they were further classified into VanR group. Mce1R, a FadR-like transcriptional regulator, is known to be involved in the regulation of virulence (Casali et al., 2006) via affecting cholesterol transport and metabolism in bacteria (Mohn et al., 2008; Miner et al., 2009). This result provides a clue that XAC0568 and XAC0877 may regulate virulence in pathogenic process. However, to elucidate this assumption, a future detailed study is required.

YtrA-like regulator of Xac

The YtrA subfamily is the least represented GntR-like regulator in the bacterial genomes (Rigali et al., 2002), which is also evident in case of *Xac*. Among seven *Xac* GntR regulators, only one regulator XAC1548 showed the signatures of the YtrA subfamily member with only two α-helices (i.e., α4 and α5) (Table 1, Fig. 2b). YtrA possesses a reduced C-terminal domain with only two α-helices. The average length of the putative E-b/O domain is about 50 amino acids (Rigali et al., 2002). YtrA from *B. subtilis* has been confirmed to be a part of a large self-regulated operon and this operon consists of genes encoding the ATP binding cassette transport systems in addition to the YtrA (Yoshida et al., 2000). Whether XAC1548 has any role in modulating such an operon seems worthy of further study.

MocR-like regulator of Xac

Only XAC0737 in the seven putative *Xac* GntR-like regulators was classified into the MocR subfamily (Table 1). It showed distinguishable predicted secondary structural features specific to this subfamily and exceptional average of about 350 amino acids (Rigali et al., 2002). MocR-like regulators show homology to class I aminotransferase proteins (Sung et al., 1991), which require pyridoxal 5'-phosphate (PLP) as a co-factor. All MocR-like regulators present a PLP attachment site with a conserved lysine residue, which is also confirmed as in XAC0737 (Fig. 2c). The study of pyridoxal phosphate regulation role in XAC0737 would be interesting.

HutC-like regulators of Xac

Contrary to the FadR-like regulators, the regulators of HutC-like subfamily consist of α+β type (i.e., α4-α9 and β3-β9) structures in the C-terminal domain. Three GntRs (XAC1640, XAC0711 and XAC3532) as members of this subfamily were identified (Table 1). These three HutC members showed subfamily-specific secondary structural features (Fig. 2d) (Rigali et al., 2002). HutC regulators are known to acquire the same protein fold as *Escherichia coli* UbiC. Consequently, it is also named as UbiC transcription regulator-associated domain (Aravind and Anantharaman, 2003). This effector-binding domain responds to various ligands (Allison and Phillips, 1990; Quail et al., 1994; Matthijs et al., 2000). A range of known ligands, specific to many HutC-like regulators, will help in characterizing the classified *Xac* regulators.

3D structure prediction and assessment of* Xac *GntRs

We were interested in developing 3D models of all the putative *Xac* GntR-like regulators in order to gain knowledge about the precise information of how proteins interact and localize in their stable conformation. It is an effective way to obtain useful information about the proteins of interest. 3D structure models of seven *Xac* GntR-like regulators were constructed using I-TASSER (Roy et al., 2010). Following the I-TASSER procedures, 3D structures were generated (Fig. 3), and the results of the backbone conformations were evaluated using the 'Verify 3D', 'PROCHECK' (Laskowski et al., 1993) and 'WHAT CHECK' programs from the 'WHAT IF' suite (Vriend, 1990). The backbone conformation of the modeled structure was calculated by analyzing the phi (φ) and psi (ψ) torsion angles using PROCHECK, as determined by Ramachandran plot statistics. The φ/ψ torsion angles of the backbone conformation of each residue are represented by Ramachandran plot. Over 95.4% of the residues were in either the favored regions or the allowed regions (Table 2, Fig. 3), which denote stable structures. Moreover, these 3D structure features of the models (Fig. 3) are consistent with the results of their secondary structure analysis (Fig. 2) as well. The ProQ neural network provided an LGscores for the model and the template structure, respectively (Table 2). This method considers that models with LGscore>1.5 are fairly good, LGscore>2.5 are very good (Cristobal et al., 2001). Fig. 3 shows the Verify 3D profiles calculated for the template and the homology model. Residues with an averaged 3D-1D score over 0.2 should be considered reliable. The results showed that 94.17, 84.89, 92.19, 67.77, 78.59, 77.82 and 80.66% of the residues have a score of over 0.2 in XAC0568, XAC0877, XAC0711, XAC1548, XAC0737, XAC1640 and XAC3532 models, respectively, indicating good models. The ProSA-web server quality assessments provide Z-score of the template and model structure, respectively (Supplementary Fig. S1). These results showed that Z-scores in the these models were negative in most residues, further indicating that all the *Xac* GntRs models are of similar quality as equivalent sized X-ray structures. All these results indicate that our structural models of the *Xac* GntR regulators are reliable from the statistical aspect. These

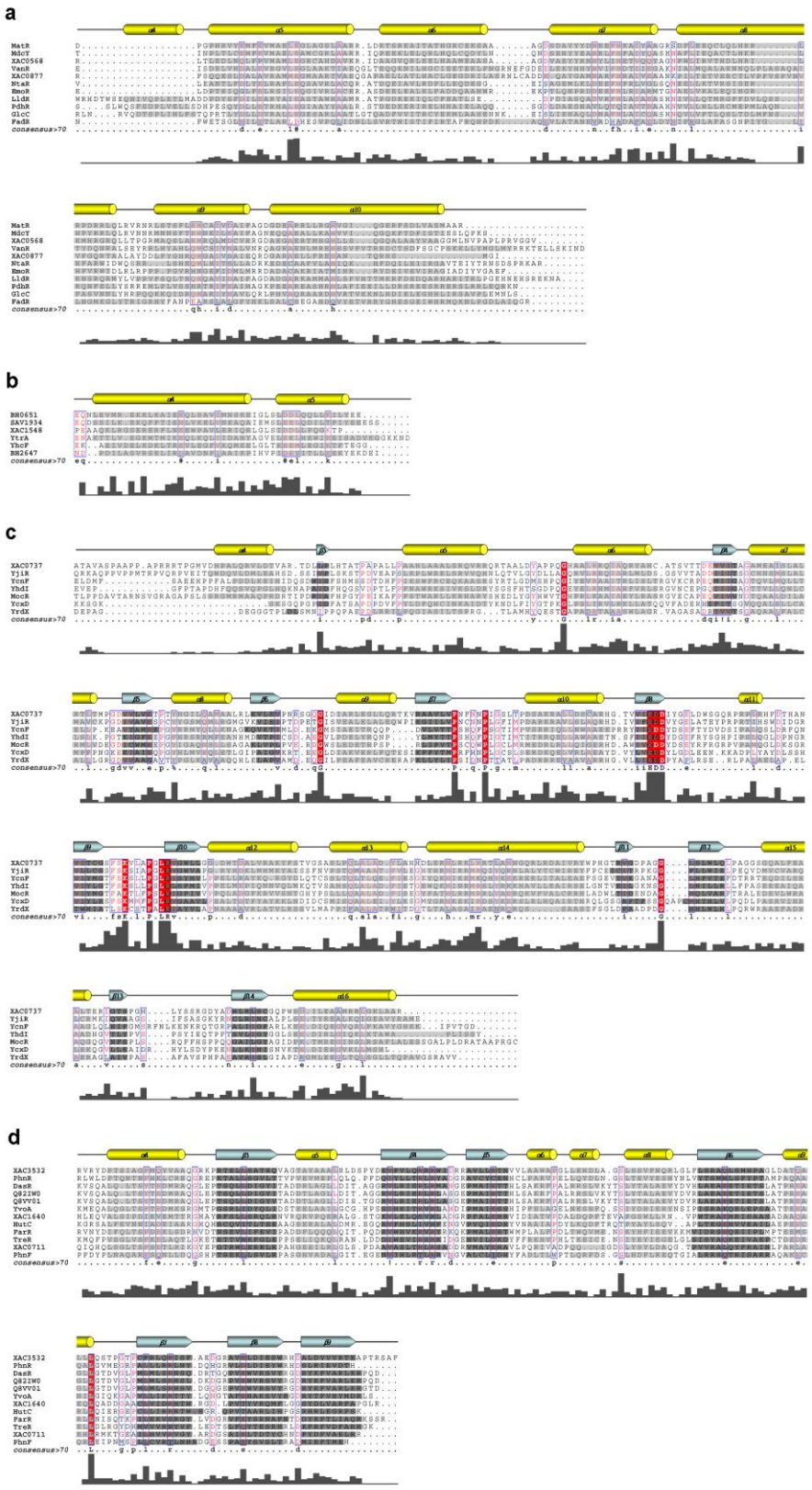

Fig 2. Structure based multiple protein sequence alignment of C-terminal of *Xac* GntR-like regulators as well as previously identified GntR regulators. The 70% consensus sequence was generated by ESPript: capital letters indicate identity, and lowercase letters indicate a consensus level of >0.5. Consensus symbol # is anyone of NDQEBZ; ! used for anyone of IV; % is anyone of FY. In graphical representation α-helix and β-sheet regions are highlighted with light and dark gray background, respectively. **(a)** FadR subfamily. **(b)** YtrA subfamily. **(c)** MocR subfamily. **(d)** HutC subfamily. Figure prepared using ESPript.

homology models are suitable for further analysis and studies.

Prediction binding sites of* Xac *GntRs

After the final model was built, the possible binding sites of the modeled structures were searched using the BSpred online program. Ranking of the predicted binding sites is based on the size of the clusters formed after local superposition of ligands of the templates onto the query structure. The best ranked possible binding site residues in models of *Xac* GntR-like regulators were showed in Fig. 4 and Table 3. An organic acid, 4-[3-(2-chloro-4,5-difluoro-benzoyl) ureido]-3-trifluoromethoxybenzoic acid, was found to interact with XAC0568 (Fig. 4a). This result showed that XAC0568 may play a role in organic acids catabolism, while, *β*-D-glucose and *α*-D-mannose were found to interact with XAC0877 and XAC0711, respectively (Fig. 4b and 4d). These monosaccharides are essential molecules for all forms of life. Either of them is in addition a preferred carbon source for many microorganisms as it provides carbon. These results illustrated that XAC0877 and XAC0711 may involve in regulation of carbon metabolism. In the case of XAC1548, the calcium ion was located at the core of the helical bundle with a coordination number of four (I^{12}, Y^{13}, T^{51} and R^{54}) (Fig. 4c). This calcium ion interaction network was involved in three helices of the helical bundle, which presumably assists in stabilizing the structure. The present structure first revealed that XAC1548, belongs to the GntR family, has an ability to bind a calcium ion, although the detailed biofunction of the calcium ion is not yet clear. XAC1548 may be functional special in the YtrA subfamily for this structural feature. While, in the case of XAC1640, the sulfate ion, a ligand attaching either by one oxygen (monodentate) or by two oxygens as either a chelate or a bridge, was found to present in the binding pocket of the 3D modeled structure with a coordination number of seven (D^{80}, E^{82}, R^{93}, S^{125}, V^{157}, A^{158} and P^{159}) (Fig. 4e). This interaction network was involved in six helices of the helical bundle, which presumably assists in stabilizing the structure. The present structure revealed that XAC1640 may take part in chemical catalysis. An aminocoumarin antibiotic, clorobiocin (CBN) ligand, was found to act as an inhibitor complexed with XAC3532 (Fig. 4f). This result showed that XAC3532, a putative repressor, complexed with clorobiocin, may play roles in 2-aminoethylphosphonate uptake and metabolism. XAC0737, a member of the MocR subfamily of transcriptional regulators contains a class I aminotransferase domain binding PLP (Fig. 2g and 4g). Like other aminotransferases, XAC0737 might transaminate taurine through conversion of PLP to pyridoxamine 5'-phosphate (PMP). In turn, XAC0737 (carrying PMP) might activate its target gene transcription. Since the putative PLP binding site is highly conserved in MocR subfamily (data not shown), we analyzed XAC0737-mediated gel retardation of its target gene promoter in the presence and absence of PLP. Along with our structure based prediction, these data suggest that *Xac* GntR regulators may play key roles in many and varied primary metabolic processes, responding to changing metabolite concentrations to regulate related genes expression. Our hypothesis is well in accordance with other published statements (Mohn et al., 2008; Hoskisson and Rigali, 2009).

Ortholog prediction

A number of orthologous proteins of putative *Xac* GntRs were found in the other species of xanthomonads (Table 4). Since orthologs typically share the similar function, the advanced research on these seven putative GntR regulators could serve as a model to study homologues from the other species of xanthomonads. These characterized orthologs may also provide clues for initiating detailed biochemical and molecular characterization of *Xac* proteins. Some putative orthologs were experimentally confirmed. For example, XC_2736 (UniProt ID: Q4UT38), a high homology ortholog (98%) to XAC1548, is involved in regulation of virulence and hypersensitive response (An et al., 2011). Otherwise, the *XAC1548* putative promoter region contains an imperfect plant-inducible promoter (PIP) box [TTCGC-N20-TTCGC; consensus TTCGC-N15-TTCGC (Wengelnik and Bonas, 1996)] 28 bp upstream of the translation initiation codon ATG. PIP box has already been found upstream of a large number of pathogenicity or virulence genes, such as *hrp* pathogenicity operons in *Xanthomonas* spp. (Stefan and Bonas, 1995; Astua-Monge et al., 2000; Buttner and Bonas, 2002). It is presumed that XAC1548 may have an important role on virulence regulation. These results illustrated that we could construct the *XAC1548* mutant strain to confirm its proposed role in *Xac*.

Materials and methods

Selection of GntR-like regulators in the genome of Xac strain 306

Apart from classified GntR regulators or proteins annotated as GntR-like regulator, other putative GntRs from *Xac* proteome were selected using EMBL and Pfam databases (Pfam entry PF00392 GntR) (Haydon and Guest, 1991; Eddy, 1998). We also did the gene research on the NCBI website (www.ncbi.nlm.nih.gov/gene) using 'GntR' and '*Xanthomonas axonopodis*' as query keywords. Finally, for integration of these methods, we collected seven putative GntR-like proteins. All these GntR proteins were retrieved from the UniProtKB sequence database as per their Swiss-Prot ID (Table 1).

Multiple sequence alignments and phylogenetic tree construction

Alignment of selected sequences was performed using program ClustalX 2.0 (Larkin et al., 2007). The corresponding phylogenetic tree was constructed with Molecular Evolutionary Genetics Analysis version 5 (MEGA5) software (Tamura et al., 2011) by employing the Neighbor Joining (NJ) method. All the parameters were taken to the default settings.

Secondary structure prediction

Multiple sequence alignment of C-terminal domains of *Xac* GntR-like regulators and other known and classified GntR regulators was performed using ClustalX 2.0 with the default parameters. Secondary structure predictions were conducted using NetSurfP (Petersen et al., 2009), Jpred (Cuff et al., 1998) and 3D-PSSM (Kelley et al., 2000). A consensus of all the secondary structure predictions was considered for a more accurate outcome.

Table 3. List of predicted binding site residues of GntR-like regulators from *Xac*.

GntR-like regulator	Binding site residues[#]
XAC0568	L^{48}, G^{49}, I^{50}, S^{51}, P^{54}
XAC0877	S^{43}, R^{44}, T^{45}, P^{46}, Q^{49}, A^{50}
XAC0711	V^{73}, D^{74}, V^{77}, L^{81}, T^{90}, F^{91}
XAC1640	D^{80}, E^{82}, R^{93}, S^{125}, V^{157}, A^{158}, P^{159}
XAC3532	A^{12}, A^{15}, Q^{16}, S^{18}, E^{21}, R^{27}, L^{28}, P^{29}, L^{34}, L^{52}, R^{60}, W^{66}, F^{67}
XAC1548	I^{12}, Y^{13}, T^{51}, R^{54}
XAC0737	A^{176}, M^{177}, Y^{201}, V^{246}, N^{251}, D^{279}, I^{281}, Y^{282}, S^{309}, S^{311}, K^{312}, R^{319}

[#] The amino acids are represented using the single letter code and the positions are indicated with superscript numbers.

Table 4. Orthologs of GntR-like regulators from other bacterial species in *Xanthomonas*.

Xac	XAC0568	XAC0711	XAC0737	XAC0877	XAC1548	XAC1640	XAC3532
Xal	-	XALc_0837	XALc_1932	XALc_3035	XALc_1658	XALc_1714	-
Xcb	XC_0580	XC_0750	XC_0778	XC_3428	XC_2736	XC_2652	XC_3561
Xcc	XCC3617	XCC3414	XCC3386	XCC0803	XCC1500	XCC1582	XCC0672
Xca	xccb100_0597	xccb100_0783	xccb100_0813	xccb100_3548	xccb100_2773	xccb100_2678	xccb100_3683
Xcv	XCV0602	XCV0767	-	XCV0912	XCV1591	XCV1682	XCV3656
Xoo	-	XOO3920	-	XOO3735	XOO2063	XOO2395	XOO0859
Xom	-	XOO_3699	-	-	XOO_1943	XOO_2275	XOO_0783
Xop	-	PXO_04320	-	PXO_04632	PXO_01221	PXO_00724	PXO_02688

'-' Corresponding orthologs are not present in the genome. Species abbreviations: *X. lbilineans* (*Xal*), *X. campestris* pv. *campestris* strain 8004 (*Xcb*), *X. campestris* pv. *campestris* strain ATCC 33913 (*Xcc*), *X. campestris* pv. *campestris* strain B100 (*Xca*), *X. campestris* pv. *vesicatoria* strain 85-10 (*Xcv*), *X. oryzae* pv. *oryzae* KACC10331 (*Xoo*), *X. oryzae* pv. *oryzae* MAFF 311018 (*Xom*), and *X. oryzae* pv. *oryzae* PXO99A (*Xop*).

Structural modeling and evaluation of Xac GntRs

Generation of full tertiary structure models was carried out through the iterative threading assembly refinement (I-TASSER) server (Roy et al., 2010) (http://zhanglab.ccmb.med.umich.edu/I-TASSER). The I-TASSER process is a combination of threading, reduced model structural fabrication, selection, adjustment, and functional determination. The I-TASSER server returns the best five models with a c-score attached for each model. Also it returns the top ten templates used in the threading. The c-score is a confidence score for estimating the quality of the predicted models by I-TASSER. The calculation of C-score is based on the significance of the threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5 to 2], where a C-score, a higher value of model with high confidence and vice-versa (Zhang, 2008). When selecting a template for modeling, we select the model that comes from the largest cluster and has the best C-score, and the corresponding template PDB with its chain (Table 2). The developed model was further refined by calculation of free energy of the system and further minimized via GROMOS96 software, incorporated in Swiss-PdbViewer v4.0.4 program (Guex and Peitsch, 1997). Finally, the refined models were verified using PROCHECK program (Laskowski et al., 1993), 3-Dimensional Structural Superposition (3d-SS) tool (Sumathi et al., 2006), and Verify 3D server (Bowie et al., 1991; Luthy et al., 1992). RAMPAGE web server was used to prepare Ramachandran plots (Lovell et al. 2003). Additional quality model assessments were performed using the ProSA-web (Sippl, 1993; Wiederstein and Sippl, 2007), ProQ (Wallner and Elofsson, 2003) and SAVES (http://nihserver.mbi.ucla.edu/SAVES) servers. Finally, all evaluated models were deposited into Protein Model DataBase (http://mi.caspur.it/PMDB). The PMDB IDs of the submissions are shown in Table 2.

Binding sites prediction

In order to assess the binding sites of the modeled structures, we submitted the primary amino acid sequences of *Xac* GntR regulators to the BSpred online program available at http://zhanglab.ccmb.med.umich.edu/BSpred, which is a neural network based algorithm for predicting binding site of proteins from amino acid sequences (Mukherjee and Zhang, 2011). The figures were prepared using Molecular Dynamics (VMD) v1.9.1 software (William et al., 1996).

Reciprocal BLAST

Reciprocal BLAST hits have been frequently utilized to identify the orthologs in two species (Fulton et al., 2006). With this method, we searched for the best reciprocal BLAST hit for *Xac* GntR proteins with *X. albilineans*, *X. campestris* pv. *campestris* strain 8004, *X. campestris* pv. *campestris* strain ATCC 33913, *X. campestris* pv. *campestris* strain B100, *X. campestris* pv. *vesicatoria* strain 85-10, *X. oryzae* pv. *oryzae* KACC10331, *X. oryzae* pv. *oryzae* MAFF 311018 and *X. oryzae* pv. *oryzae* PXO99A.

Conclusion

We classified seven putative *Xac* GntRs into four subfamilies FadR, HutC, MocR and YtrA according to their C-terminal secondary structure features. Then, the homology models were predicted based on the secondary and tertiary structure information. We suggest that this is an effective way to obtain useful information about the *Xac* GntRs. Potential binding sites were also ascertained by analyzing the primary amino acid sequences of GntR-like regulators. This finding may provide valuable clues for initiating the experimental characterization of these regulators and

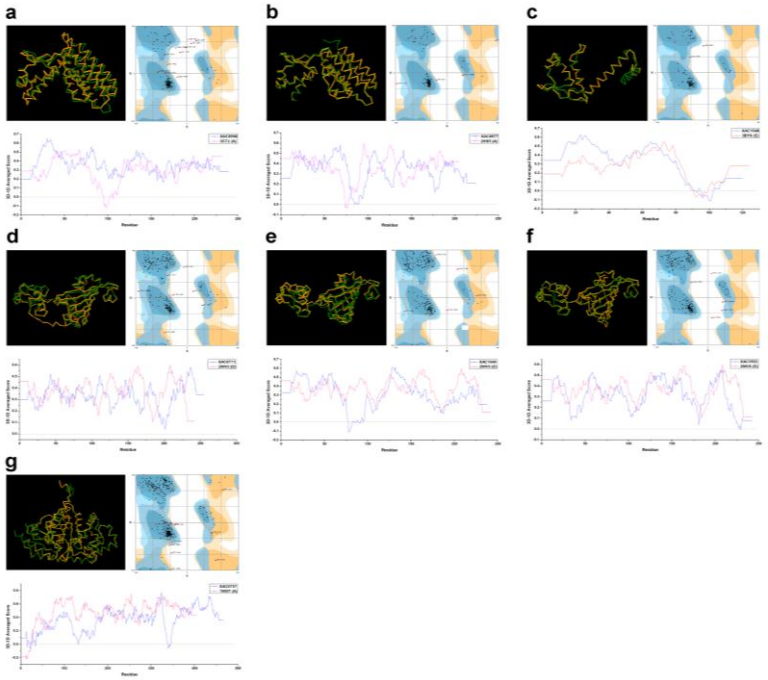

Fig 3. Comparative modeling of *Xac* GntR-like regulators. **(a)** XAC0568, **(b)** XAC0877, **(c)** XAC1548, **(d)** XAC0711, **(e)** XAC1640, **(f)** XAC3532, **(g)** XAC0737. **(*Upper left panel*)** Structural superposition of *Xac* GntR homology molecular models (*yellow*) on their corresponding templates (*green*) uses the 3d-SS tool, revealing the overall similarity in backbone conformations. **(*Upper right panel*)** Ramachandran plots evaluated by the RAMPAGE program, the red areas (i.e., A, B and L) correspond to the most favored regions; the yellow areas (i.e., a, b, l and p) refer to additionally allowed regions; the light yellow areas (i.e., ~a, ~b, ~l and ~p) represent the generously allowed regions and the white areas stand for disallowed regions. **(*Lower panel*)** Verify 3D score profiles for the generated models and template structures obtained using the SAVES server. Scores greater than 0.2 indicate a high-quality structure.

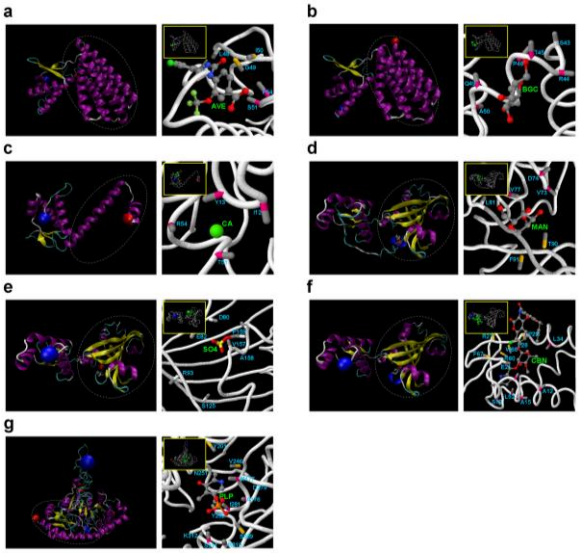

Fig 4. Topology and binding sites of *Xac* GntR-like regulators. **(a)** XAC0568, **(b)** XAC0877, **(c)** XAC1548, **(d)** XAC0711, **(e)** XAC1640, **(f)** XAC3532, **(g)** XAC0737. **(*Left panel*)** The GntR-like regulator homology models utilizing automated protein modeling option were retrieved through ESyPred3D. The structure is displayed as a solid ribbon diagram. The *α*-helices and *β*-sheets are shown as *purple helices* and *yellow ribbons*, respectively. The rest of the structure is depicted as *loops*. N- and C-terminus residues are shown as *blue* and *red* spheres, respectively, while C-terminal domains were shown by *white ellipses*. **(*Right panel*)** Upper left, the possible binding sites of GntRs from *Xac*. Binding site residues in the model were predicted by BSpred online program. Predicted binding site residues in the models are shown in green sphere, while N- and C-terminus residues are shown as blue and red spheres, respectively. In the small insert from upper left, the close-up view of the binding pocket localization is shown. The surface of the modeled protein is shown as traces, ligands as balls-and-sticks, and the highly conserved residues probably involved in the binding are indicated as sticks and colored in deepskyblue. Abbreviations: AVE, 4-[3-(2-chloro-4,5-difluoro-benzoyl) ureido]-3-trifluoromethoxybenzoic acid; BGC, *β*-D-glucose; CA, calcium ion; CBN, clorobiocin; MAN, *α*-D-mannose; PLP, pyridoxal 5'-phosphate; SO4, sulfate ion.

enhance our understanding of the influence of these proteins on the virulence or primary metabolism of citrus canker pathogen. Moreover, the orthologs prediction was also conducted by selecting best homologous proteins showing maximum sequence similarity with the unknown *Xac* GntR regulators to predict the possible function of these unknown proteins, and vice versa. The identified orthologs from *Xac* could serve as a model to decipher molecular regulation events taking place in plant-pathogen interactions. Though we have only made an attempt to explore *Xac* GntR regulators, this approach could also be effectively employed to extend the GntR family classification in other plant pathogenic microbes as well.

Acknowledgements

This work was supported by the International Science and Technology Cooperation Project funded by the Ministry of Science and Technology of P. R. China (2011DFA32030), the Special Fund for Agro-scientific Research of the Public Interest (201003067-07), and the National Science Fund for Distinguished Young Scholars of China (31000901).

References

Al-Saadi A, Reddy JD, Duan YP, Brunings AM, Yuan Q, Gabriel DW (2007) All five host-range variants of *Xanthomonas citri* carry one *pthA* homolog with 17.5 repeats that determines pathogenicity on citrus, but none determine host-range variation. Mol Plant-Microbe Interact 20: 934-943

Allison S, Phillips A (1990) Nucleotide sequence of the gene encoding the repressor for the histidine utilization genes of *Pseudomonas putida*. J Bacteriol 172: 5470-5476

An S, Lu G, Su H, Li R, He Y, Jiang B, Tang D, Tang J (2011) Systematic mutagenesis of all predicted *gntR* genes in *Xanthomonas campestris* pv. *campestris* reveals a GntR family transcriptional regulator controlling hypersensitive response and virulence. Mol Plant-Microbe Interact 24: 1027-1039

Aravind L, Anantharaman V (2003) HutC/FarR-like bacterial transcription factors of the GntR family contain a small molecule-binding domain of the chorismate lyase fold. FEMS Microbiol Lett 222: 17-23

Astua-Monge G, Minsavage GV, Stall RE, Davis MJ, Bonas U, Jones JB (2000) Resistance of tomato and pepper to T3 strains of *Xanthomonas campestris* pv. *vesicatoria* is specified by a plant-inducible avirulence gene. Mol Plant-Microbe Interact 13: 911-921

Bowie J, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253: 164-170

Buttner D, Bonas U (2002) Getting across-bacterial type III effector proteins on their way to the plant cell. EMBO J 21: 5313-5322

Casali N, White AM, Riley LW (2006) Regulation of the *Mycobacterium tuberculosis mce1* operon. J Bacteriol 188: 441-449

Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) JPred: a consensus secondary structure prediction server. Bioinformatics 14: 892-893

Das AK (2003) Citrus canker - A review. J Appl Hort 5: 52-60

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14: 755-763

Frunzke J, Engels V, Hasenbein S, Gätgens C, Bott M (2008) Co-ordinated regulation of gluconate catabolism and glucose uptake in *Corynebacterium glutamicum* by two functionally equivalent transcriptional regulators, GntR1 and GntR2. Mol Microbiol 67: 305-322

Fulton DL, Li YY, Laird MR, Horsman BG, Roche FM, Brinkman FS (2006) Improving the specificity of high-throughput ortholog prediction. BMC Bioinformatics 7: 270

Graham JH, Gottwald TR, Cubero J, Achor DS (2004) *Xanthomonas axonopodis* pv. *citri*: factors affecting successful eradication of citrus canker. Mol Plant Pathol 5: 1-15

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis 18: 2714-2723

Haine V, Sinon A, Van Steen F, Rousseau S, Dozot M, Lestrate P, Lambert C, Letesson J-J, De Bolle X (2005) Systematic targeted mutagenesis of *Brucella melitensis* 16M reveals a major role for GntR regulators in the control of virulence. Infect Immun 73: 5578-5586

Haydon DJ, Guest JR (1991) A new family of bacterial regulatory proteins. FEMS Microbiol Lett 79: 291-296

Hoskisson PA, Rigali S (2009) Chapter 1 Variation in form and function: The helix-turn-helix regulators of the GntR superfamily. In: Allen IL, Sima S, Geoffrey MG (eds) Adv. Appl. Microbiol., Volume 69. Academic Press, pp 1-22

Hyeon JE, Kang DH, Kim YI, You SK, Han SO (2012) A GntR type transcriptional regulator PckR negatively regulates the expression of phosphoenolpyruvate carboxykinase in *Corynebacterium glutamicum*. J Bacteriol. 194 (9):2181-2188

Ji L, Xie JP (2011) GntR family regulators of the pathogen of fish tuberculosis *Mycobacterium marinum*. Biochem Bioph Res Co 410: 780-785

Kelley L, MacCallum R, Sternberg M (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol 299: 499-520

Laia M, Moreira L, Dezajacomo J, Brigati J, Ferreira C, Ferro M, Silva A, Ferro J, Oliveira J (2009) New genes of *Xanthomonas citri* subsp. *citri* involved in pathogenesis and adaptation revealed by a transposon-based mutant library. BMC Microbiol 9: 12

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26: 283-291

Lee M, Scherer M, Rigali S, Golden J (2003a) PlmA, a new member of the GntR family, has plasmid maintenance functions in *Anabaena* sp. strain PCC 7120. J Bacteriol 185: 4315-4325

Li JY, Wang N (2012) The *gpsX* gene encoding a glycosyltransferase is important for polysaccharide production and required for full virulence in *Xanthomonas citri* subsp. *citri*. BMC Microbiol 12: 31

Liu LP, Deng ZN, Qu JW, Yan JW, Catara V, Li DZ, Long GY, & Li N (2012) Construction of EGFP-labeling system for visualizing the infection process of *Xanthomonas axonopodis* pv. *citri in planta*. Curr Microbiol 65 (3) 304-312.

Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Cα geometry: phi, psi and Cβ deviation. Proteins Struct Funct Bioinf 50: 437-450

Lu H, Patil P, Van Sluys M-A, White FF, Ryan RP, Dow JM, Rabinowicz P, Salzberg SL, Leach JE, Sonti R, Brendel V, Bogdanove AJ (2008) Acquisition and evolution of plant pathogenesis–associated gene clusters and candidate determinants of tissue-specificity in *Xanthomonas*. PLoS One 3: e3828

Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356: 83-85

Malamud F, Torres PS, Roeschlin R, Rigano LA, Enrique R, Bonomi HR, Castagnaro AP, Marano MR, Vojnov AA (2011) The *Xanthomonas axonopodis* pv. *citri* flagellum is required for mature biofilm and canker development. Microbiol 157: 819-829

Matthijs S, Koedam N, Cornelis P, De Greve H (2000) The trehalose operon of *Pseudomonas fluorescens* ATCC 17400. Res Microbiol 151: 845-851

Miner MD, Chang JC, Pandey AK, Sassetti CM, Sherman DR (2009) Role of cholesterol in *Mycobacterium tuberculosis* infection. Indian J Exp Biol 47: 407-411

Mohn WW, van der Geize R, Stewart GR, Okamoto S, Liu J, Dijkhuizen L, Eltis LD (2008) The actinobacterial *mce4* locus encodes a steroid transporter. J Biol Chem 283: 35368-35374

Mukherjee S, Zhang Y (2011) Protein-protein complex structure predictions by multimeric threading and template recombination. Structure 19: 955-966

Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C (2009) A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct Biol 9: 51

Quail MA, Dempsey CE, Guest JR (1994) Identification of a fatty acyl responsive regulator (FarR) in *Escherichia coli*. FEBS Lett 356: 183-187

Rigali S, Derouaux A, Giannotta F, Dusart J (2002) Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. J Biol Chem 277: 12507-12515

Rigali S, Schlicht M, Hoskisson P, Nothaft H, Merzbacher M, Joris B, Titgemeyer F (2004) Extending the classification of bacterial transcription factors beyond the helix-turn-helix motif as an alternative approach to discover new *cis*/*trans* relationships. Nucleic Acids Res 32: 3418-3426

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5: 725-738

Ryan RP, Vorhölter F-J, Potnis N, Jones JB, Van Sluys M-A, Bogdanove AJ, Dow JM (2011) Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. Nat Rev Microbiol 9: 344-355

da Silva ACR, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorello CB, Sluys MAV, Almeida NF, Alves LMC, do Amaral AM, Bertolini MC, Camargo LEA, Camarotte G, Cannavan F, Cardozo J, Chambergo F, Ciapina LP, Cicarelli RMB, Coutinho LL, Cursino-Santos JR, El-Dorry H, Faria JB, Ferreira AJS, Ferreira RCC, Ferro MIT, Formighieri EF, Franco MC, Greggio CC, Gruber A, Katsuyama AM, Kishi LT, Leite RP, Lemos EGM, Lemos MVF, Locali EC, Machado MA, Madeira AMBN, Martinez-Rossi NM, Martins EC, Meidanis J, Menck CFM, Miyaki CY, Moon DH, Moreira LM, Novo MTM, Okura VK, Oliveira MC, Oliveira VR, Pereira HA, Rossi A, Sena JAD, Silva C, de Souza RF, Spinola LAF, Takita MA, Tamura RE, Teixeira EC, Tezza RID, Trindade dos Santos M, Truffi D, Tsai SM, White FF, Setubal JC, Kitajima JP (2002) Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. Nature 417: 459-463

Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. Proteins Struct Funct Genet 17: 355-362

Stefan F, Bonas U (1995) Sequence and expression analysis of the *hrpB* pathogenicity operon of *Xanthomonas campestris* pv. *vesicatoria* which encodes eight proteins with similarity to components of the Hrp, Ysc, Spa, and Fli secretion systems. Mol Plant-Microbe Interact 8: 845-854

Sumathi K, Ananthalakshmi P, Roshan MNAM, Sekar K (2006) 3dSS: 3D structural superposition. Nucleic Acids Res 34: W128-W132

Sung M, Tanizawa K, Tanaka H, Kuramitsu S, Kagamiyama H, Hirotsu K, Okamoto A, Higuchi T, Soda K (1991) Thermostable aspartate aminotransferase from a thermophilic *Bacillus* species. Gene cloning, sequence determination, and preliminary x-ray characterization. J Biol Chem 266: 2567-2572

Swarup S, De Feyter R, Brlansky RH, Gabriel DW (1991) A pathogenicity locus from *Xanthomonas citri* enables strains from several pathovars of *X. campestris* to elicit cankerlike lesions on citrus. Phytopathology 81: 802-809

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739

Vindal V, Ranjan S, Ranjan A (2007a) *In silico* analysis and characterization of GntR family of regulators from *Mycobacterium tuberculosis*. Tuberculosis 87: 242-247

Vindal V, Suma K, Ranjan A (2007b) GntR family of regulators in *Mycobacterium smegmatis*: a sequence and structure based characterization. BMC Genomics 8: 289

Vriend G (1990) WHAT IF: A molecular modeling and drug design program. J Mol Graph 8: 52-56

Wallner B, Elofsson A (2003) Can correct protein models be identified? Protein Sci 12: 1073-1086

Wengelnik K, Bonas U (1996) HrpXv, an AraC-type regulator, activates expression of five of the six loci in the *hrp* cluster of *Xanthomonas campestris* pv. *vesicatoria*. J Bacteriol 178: 3462-3469

Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35: W407-W410

William H, Andrew D, Klaus S (1996) VMD: visual molecular dynamics. J Mol Graph 14: 33-38

Yoshida K, Fujita Y, Ehrlich S (2000) An operon for a putative ATP-binding cassette transport system involved in acetoin utilization of *Bacillus subtilis*. J Bacteriol 182: 5454-5461

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9: 40