Comparative Genomics Insight into Phytopathogenic Xanthomonas arboricola pathovar corylina Strains

Şafak KALINDAMAR
Department of Molecular Biology and Genetics, Faculty of Arts and Sciences, Ordu University, 52200, Ordu, Turkey

Abstract: Xanthomonas arboricola pathovar corylina (Xac) causes a bacterial blight disease (BBD) resulting in economic losses in young hazelnut trees worldwide. Although virulent Xac genomes were sequenced, there is no comparative genomics study on these genomes. In this study, all publicly available whole-genome sequences of Xac were compared by a comparative genomics approach. The results showed that Xac CFBP1159 and Xac CFBP2565 genomes are phylogenetically related to each other based on the orthology results. The genomic diversification of Xac strains was depended on mobile genetic elements. Interestingly, Xac NCCB100457 genome had additional motility genes than Xac CFBP1159 and Xac CFBP2565 genomes. All Xac genomes had shared virulence-related genes such as secretion systems and adherence factors. The comparative genomics approach of this study supports that Xac genomes have slight genetic variations, and the virulence-related proteins interacted with the host proteins. This comparative genomics approach will provide insights into the understanding of the Xac genomes.

Keywords: Comparative genomics, host-pathogen interaction, phytopathogenic, virulence, Xanthomonas.

Fitopatojenik Xanthomonas arboricola patovar corylina Suşlarına Karşılaştırmalı Genomik Bakış

Öz: Xanthomonas arboricola patovar corylina (Xac), Dünya çapında genç fındık ağaçlarında ekonomik kayıplara sebep olan önemli bir hastalığa neden olmaktadır. Virülen Xac suşlarının tüm genomları dizilenmiş olmasına rağmen, bu genomlar üzerinde bir karşılaştırmalı genom çalışması yoktur. Bu çalışma, Xac suşlarının halka açık tüm genom dizilerini, karşılaştırmalı genomik analizler kullanarak karşılaştırmıştır. Elde edilen sonuçlarda, Xac CFBP1159 ve Xac CFBP2565 genomları, ortoloji sonucuna göre filogenetik olarak birbirlerine oldukça yakın ilişkili olduğu saptanmıştır. Xac suşlarının genomik çeşitliliğinin mobil genetik elementlerle ilgisi olduğu anlaşılmıştır. İlgincir bir şekilde, Xac NCCB100457 genomonu, Xac CFBP1159 ve Xac CFBP2565 genomlarından daha fazla motilitelik genlerine sahiptir. Tüm Xac genomları, sekresyon sistemleri ve adhezyon faktörleri gibi virülsen le ilgili örtük genler şahit edilmiştir. Bu çalışmanın karşılaştırmalı genomik analizi, Xac genomlarının bazı genetik varyasyonlarına sahip olduğuna ve virülsen le ilgili proteinerlerin konak proteinlerle etkileşime girdiği desteklemektedir. Bu çalışmanın karşılaştırmalı genomik analizi, Xac genomlarının anlaşılmasını için bir öngörü sağlamıştır.

Anahtar kelimeler: Bitki patojeni, karşılaştırmalı genomik, konak patojenilik, virülsen, Xanthomonas.
INTRODUCTION

Xanthomonas arboricola is primarily known as a Gram-negative phytopathogenic bacterium that can infect economically important plant species such as prunus, walnut, and hazelnut (Vauterin et al., 1995). X. arboricola pathovar-assigned strains are mostly described as pathogenic bacteria although there are X. arboricola strains considered non-pathogenic bacteria. Moreover, there are nine pathovar strains (arracaciae, celebensis, corylina, fragariae, guizotiae, juglandis, populi, pruni, and zanteschiae) identified in X. arboricola (Fischer-Le Saux et al., 2015). As one of the most virulent pathovars, Xanthomonas arboricola pathovar corylina (Xac) is a causative agent of BBD of hazelnut trees (Corylus spp.). The BBD was first described on C. maxima in the USA (Barss, 1913). Later, the disease has been increasingly reported on C. avellana from hazelnut producer countries in the European Union (Eppo, 2004). Xac has been also prevalently isolated from the hazelnut orchards in the Black Sea Region of Turkey, which is a major hazelnut producer in Europe (Karahan et al., 2013). The symptoms of BBD can appear on hazelnut tree’s leaves and twigs. It has been also reported that the disease can also kill hazelnut trees or delay the growth of trees (Lamicchane et al., 2014).

The genus of Xanthomonas include a variety of plant pathogenic bacterial species, and these bacteria can utilize diverse virulence mechanisms such as secretion systems, effector proteins of diverse secretion systems, flagella, and small RNAs (Timilsina et al., 2020). Recently, a comparative genomics study on virulent and non-virulent Xanthomonas arboricola pv. pruni (Xap) strains showed genomic differences between virulent and non-virulent Xap strains (Garita et al., 2017). On the other hand, the Xac strain was isolated from the leaf spots of C. corylina L. in Colorado, and the first DNA sequence of Xac was revealed in 2013 (Caballero et al., 2013). There are also sequences of two Xac genomes (Xac CFBP1159 and Xac CFBP2565) publicly stored in the National Center for Biotechnology Information (NCBI). However, these genomes were not used for any comparative genomics analysis.

Although several Xac genomes have been sequenced and their genomes were publicly available, there is no comparative genomics study on Xac’s genomes. Thus, the goal of this study was to compare Xac genomes that are publicly available in the NCBI database. This study provides a comparative insight into the comparison of Xac genomes, genomic variations, orthologous clusters at the protein level, annotation and gene predictions, prediction of virulence factors, and interactions of host-pathogen proteins.

MATERIAL AND METHOD

Bacterial genome data: The Xac genomes data were obtained from NCBI (as of 01/11/2020). The Xac genomes were listed with features in Table 1.

Comparative genome analyses: Blast Ring Image Generator (BRIG) was used to visualize the comparative genome analysis of Xac strains (Alikhan et al., 2011). The orthologous clusters of Xac strains at the protein level were calculated by using OrthoVenn2 (Wang et al., 2015). Prokaryotic Genomes Automatic Annotation Pipeline (NCBI PGAAP) and Rapid Annotation using Subsystem Technology (RAST) annotation pipelines were used for the annotation and gene prediction of bacterial genomes (Angiuoli et al., 2008; Overbeek, 2014). The potential protein-protein interactions between 3 of Xac whole proteins and the complete proteins of the eukaryotic host C. avellana (European hazelnut) genome (GenBank accession #: CAJGP010000000) were determined using the Host-Pathogen Interaction Database (HPIDB) by the default upload options (Ammari et al., 2016; Kumar and Nanduri, 2010). The predicted putative virulence factors were determined by downloading the full dataset from the Virulence Factors Database (VFDB) and by uploading them to BioEdit software (Lihong et al., 2016). The local BLAST feature of BioEdit was used with a cutoff E-value of 10−50 (Hall, 1999).

RESULTS

Genome features of Xac genomes: The genomic information about Xac genomes (plant host, location, and genome-level) was summarized in Table 1. The three genomes of Xac was represented and stored in NCBI. The average genome size of Xac genomes is 5.13 million base-pairs (Mbp). The average G+C content of Xac genomes is 65.53 mol%. The NCBI genome data indicated that Xac genomes were isolated from different hazelnut species (Table 1). There are no plasmids sequences reported for Xac strains.

Table 1. Genome features of Xac strains.

| Species     | Strain     | Host       | Location     | Size (Mbp) | G+C | Level   | Accession #               | Reference   |
|-------------|------------|------------|--------------|------------|-----|---------|---------------------------|-------------|
| Xac         | CFBP2565   | C. avellana| France       | 5.05       | 65.6| Contig  | NZ_MDS0101000001.1        | INRA        |
| Xac         | CFBP1559   | C. maxima  | USA          | 5.11       | 65.5| Scaffold| NZ_MDA010000001.1         | INRA        |
| Xac         | NCCB100457 | C. corylina| USA          | 5.23       | 65.5| Contig  | NZ_AMD010000001.1         | Caballero et al., 2013 |

Location: country of origin, Level: genome assembly status, Mbp: million base pairs, Xac: Xanthomonas arboricola pathovar corylina, C. Corylus.
Comparative analyses of *Xac* genomes: The visualization of the comparison of genes by BRIG showed that most of the genes among all *Xac* genomes were conserved (Fig 1). In orthology analysis, the comparison of proteins encoded by *Xac* genomes was determined by OrthoVenn2 (Fig 2). All strains shared 3694 clusters of orthologous proteins. The unique protein clusters identified such as 1 cluster (Non-ribosomal peptide synthetase) in *Xac* CFBP1159, 5 clusters (Phage portal protein, P-type conjugative transfer protein TrbJ, DDE-type integrase/transposase) in *Xac* CFBP2565, and 5 clusters (SDR family oxidoreductase, UvrD helicase, virulence RhuM family protein, transcriptional regulator) in *Xac* NCCB100457 was unique to genomes. The annotation and categorization of genes based upon the function such as adherence, adhesion, and virulence factors detected in *Xac* genomes based on orthology. The Venn diagrams and bar charts show the numbers of unique and shared orthologous genes of each genome. Interestingly, *Xac* NCCB100457 had motility and chemotaxis genes, which other *Xac* genomes do not encode any of these genes (Fig 3). In host-pathogen interactions (HPIs), the interactions were predicted between a total of 26 proteins of *Xac* and 34 proteins of *C. avellana* (Fig 4). The virulence factors detected in *Xac* genomes were categorized based on the function such as adherence, adhesion, and secretion systems (Table 2). The pathogen proteins interacted with the host proteins are listed in Table 3.

Figure 1. Comparative circular visualization of *Xac* genomes. The reference strain is *Xac* NCCB100457. The two inner rings represent the G+C content (black) and GC-skew (green/purple). The three outside rings represent a genomics comparison between *Xac* CFBP1159 and *Xac* CFBP2565 strains and the reference *Xac* NCCB100457 strain.

![Figure 1](image)

**Table 2. Virulence factors of *Xac* strains.**

| Class          | Virulence Factor                  | CFBP1159 | CFBP2565 | NCCB100457 |
|----------------|-----------------------------------|----------|----------|------------|
| Adherence      | Type IV pili                       | +        | +        | +          |
| Adhesion       | Outer membrane protein             |          |          |            |
| Adhesion       | Autotransporter-like protein       | +        | +        | +          |
| Secretion system | Type II secretion system      | +        | +        | +          |
| Secretion system | Type III secretion system    | +        | +        | +          |

(+: Presence of virulence factor in the genome)

![Figure 2](image)

**Figure 2.** Venn diagram of protein clusters encoded by *Xac* genomes based on orthology. The Venn diagrams and bar charts show the numbers of unique and shared orthologous genes of each genome.

**Table 3. Predicted pathogen proteins in host-pathogen interactions.**

| Species          | Protein ID               | Pathogen Protein                                                                 |
|------------------|--------------------------|----------------------------------------------------------------------------------|
| *Xac* CFBP1159   | WP_014060131.1            | Diaminopimelate epimerase                                                         |
|                  | WP_022937612.1            | ParA family protein                                                              |
|                  | WP_030815062.1            | RNA-proQ(1/4) S-adenosylmethionine hbo3-adenosyltransferase isomerase            |
|                  | WP_030815650.1            | Response regulator                                                               |
|                  | WP_030815715.1            | Histidinol dehydrogenase                                                         |
|                  | WP_030815673.1            | ParA family protein                                                              |
|                  | WP_050465317.1            | Thioredoxin ToxC                                                                |
|                  | WP_05046517.1             | Mannitol-dehydrogenase family protein                                           |
|                  | WP_10641227.1             | CHASE3 domain-containing protein                                                |
|                  | WP_030816713.1            | ParA family protein                                                              |
|                  | WP_050465317.1            | Thioredoxin ToxC                                                                |
|                  | WP_0504617.1              | Mannitol-dehydrogenase family protein                                           |
|                  | WP_10466930.1             | CHASE3 domain-containing protein                                                |
|                  | WP_10466930.1             | CHASE3 domain-containing protein                                                |
|                  | WP_022937612.1            | ParA family protein                                                              |
|                  | WP_030815715.1            | Histidinol dehydrogenase                                                         |
|                  | WP_030816173.1            | ParA family protein                                                              |
|                  | WP_050465317.1            | Thioredoxin ToxC                                                                |
|                  | WP_0504617.1              | Mannitol-dehydrogenase family protein                                           |
|                  | WP_10466930.1             | CHASE3 domain-containing protein                                                |
|                  | WP_10466930.1             | CHASE3 domain-containing protein                                                |
|                  | WP_022937612.1            | ParA family protein                                                              |
|                  | WP_030815715.1            | Histidinol dehydrogenase                                                         |
|                  | WP_030816173.1            | ParA family protein                                                              |

![Figure 3](image)

*Xac: Xanthomonas arboricola pathovar corylina*
DISCUSSION AND CONCLUSION

In this research, all publicly available genomes of Xac were analyzed by comparative genomics methods. This is the first comparative genomics analysis of Xac genomes isolated from three different hazelnut species.

Orthology analysis provides more accurate information about the phylogenetic relationship between closely related bacterial strains (Ullah et al., 2015). Based on the orthology analysis of Xac genomes in this study, Xac CFBP1159 and Xac CFBP2565 genomes are more closely related than Xac NCCB100457 genome. On the other hand, Xac strains share 3694 orthologous clusters. The genetic differences among genomes depend on mobile elements. This finding suggests that genetic differentiation of Xac genomes is linked to the acquisition of different mobile elements. Interestingly, Xac NCCB100457 genome possesses additional motility and chemotaxis gene sets.

This variation may be related to location and host adaptation.

The prediction of potential interactions between host and pathogen proteins may provide valuable information about an infection process (Durmus et al., 2015). The HPIs analysis in this study showed that a total of 26 proteins of Xac proteins interacted with host C. avellana proteins. These results confirm that multiple potential interactions occur between Xac and C. avellana in the host-pathogen interactions. Thus, the HPIs result demonstrated evidence of Xac virulence during infection.

Determination of virulence factors of bacterial pathogens is important to understand the pathogenesis of bacteria during the infection process (Wu et al., 2008). The whole-genome sequencing of Xac strains enables identifying virulence-related factors. Identification of virulence factors in this study revealed that Xac genomes have important virulence-related factors such as type II secretion system (T2SS), type III secretion system (T3SS), and adhesion.
factors (Table 2). The bacterial adhesion of Xanthomonas strains is crucially important for the invasion of plant tissue during the diseases process (Mhedbi-Hajri et al., 2011). While Xac strains invade plant tissues, degradative enzymes and secretion system-dependent effector proteins play a major role in establishing a successful bacterial infection. For example, T3SS and effector proteins are identified as important virulence factors in plant immunity suppression by Xac strains (Hajri et al., 2011; Jacques et al., 2016). In addition to T3SS, T2SS also plays an important role in secreting a variety of degradation enzymes in the genus of Xanthomonas (Szczesny et al., 2010). Overall, these virulence factors may contribute to the bacterial pathogenicity of Xac strains in the host.

As a result, the comparative genomics data presented in this study showed that Xac strains are closely related to each other. The virulence-related factors of Xac strains are also important in host-pathogen interactions.

ACKNOWLEDGEMENTS

This work was supported by the Department of Molecular Biology and Genetics, Faculty of Arts and Sciences, Ordu University, Turkey.

REFERENCES

Alikhan, N.F., Petty, N.K., Ben Zakour, N.J., & Beatson, S.A. (2011). Blast ring image generator (BRIG): simple prokaryote genome comparisons. **BMC Genomics**, 12, 402.

Ammari, M.G., Gresham, C.R., McCarthy, F.M. & Nanduri, B. (2016). HPIDB 2.0: a curated database for host-pathogen interactions. **Database**, (Oxford), **2016**, baw103.

Angioli, S.V., Gussman, A., Klimke, W., Cochrane, G., Field, D., Garrity, G., Kodira, C.D., Kyrpides, N., Madupu, R., Markowitz, V., Tatusova, T., Thomson, N. & White, O. (2008). Toward an online repository of Standard Operating Procedures (SOPs) for (meta) genomic annotation. **OMICS**, 12, 137-141.

Barss, H.P. (1913). A new blight disease in Oregon. *Oregon Agricultural Experiment Station Biennial Crop Pest and Horticulture Report*, 14, 213-23.

Caballero, J.I., Zerillo, M.M., Snelling, J., Boucher, C. & Tisserat, N. (2013). Genome sequence of *Xanthomonas arboricola pv. corylina*, isolated from Turkish Filbert in Colorado. *Genome Announcement*, 1(3), e00246-13. DOI: 10.1128/genomeA.00246-13

Durmus, S., Cakir, T., Ozgur, A. & Guthke, R. (2015). A review on computational systems biology of pathogen-host interactions. *Frontiers in Microbiology*, 6, 235. DOI: 10.3389/fmicb.2015.00235

EPPO. 2004. (2004). Diagnosis protocols for regulated pests *Xanthomonas arboricola* pv. *corylina*. **EPPO Bulletin**, 34, 155-7.

Fischer-Le Saux, M., Bonneau, S., Essakhli, S., Manceau, C. & Jacques, M.A.A. (2015). Aggressive emerging pathovars of *Xanthomonas arboricola* represent widespread epidemic clones distinct from poorly pathogenic strains, as revealed by multilocus sequence typing. *Applied Environmental Microbiology*, 81, 4651-4668. https://doi.org/10.1128/AEM.00050-15

Garita, C.I., Palacio-Bielsa, A., López, M.M. & Cubero, J. (2017). Pan-Genomic analysis permits differentiation of virulent and non-virulent Strains of *Xanthomonas arboricola* that Cohabit *Prunus* spp. and elucidate bacterial virulence factors. *Frontiers in Microbiology*, 8, 573. DOI: 10.3389/fmicb.2017.00573

Hajri, A., Pothier, J.F., Saux, M.F.L. Bonneau, S., Poussier, S., Bouroue, T., Duffy, B. & Manceau, C. (2011). Type three effector gene distribution and sequence analysis provide new insights into the pathogenicity of plant-pathogenic *Xanthomonas arboricola*. *Applied and Environmental Microbiology*, 78(2), 371-384. DOI: 10.1128/AEM.06119-11

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95-98.

Jacques, M.A., Arlat, M., Boulanger, A., Bouroue, T., Carrière, S., Cesbron, S., Chen, N.W., Cociancich, S., Darrasse, A., Denancé, N., Fischer-Le Saux, M., Gagnevin, L., Koebnik, R., Lauber, E., Noël, I.D., Pieretti, I., Portier, P., Pruvost, O., Rieux, A., Robèine, I., Royer, M., Szurek, B., Verdirer, V. & Vernière, C. (2016). Using Ecology, Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*. *Annual Reviews Phytopathology*, 54, 163-87. DOI: 10.1146/annurev-phyto-080615-100147

Karahan, A., Altundağ, Ş., Duran, H. & Kılıç, A.O. (2013). Karadeniz Bölgesinde fındık bakteriyel yankıları [Xanthomonas arboricola pv. corylina (Miller et al.) Vauterin et al.] hastalığını yaygınlığı üzerine araştırmalar. *Bitki Koruma Bülteni*, 53(3), 159-174.

Kumar, R. & Nanduri, B. (2010). HPIDB - a unified resource for host-pathogen interactions. **BMC Bioinformatics**, 11(6), S16. DOI: 10.1186/1471-2105-11-S6-S16

Lamichhane, J.R. & Varvaro, L. (2014). *Xanthomonas arboricola* disease of hazelnut: current status and future perspectives for its management. *Plant Pathology*, 63(2), 243-254. DOI: 10.1111/ppa.12152

Lihong, C., Dandan, Z., Bo, L., Jian, Y. & Qi, J. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis-10 years on. *Nucleic Acids
Mhedbi-Hajri, N., Darrasse, A., Pigne, S., Durand, K., Fouteau, S., Barbe, V., Manceau, C., Lemaire, C. & Jacques, M.A. (2011). Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads. *BMC Evolutionary Biology, 11*, 67.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F. & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research, 42*, D206-D214. DOI: 10.1093/nar/gkt1226

Szczesny, R., Jordan, M., Schramm, C., Schulz, S., Cogez, V., Bonas, U. & Büttner, D. (2010). Functional characterization of the Xcs and Xps type II secretion systems from the plant pathogenic bacterium *Xanthomonas campestris pv vesicatoria*. *New Phytologist, 187*(4), 983-1002. DOI: 10.1111/j.1469-8137.2010.03312.x

Timilsina, S., Potnis, N., Newberry, E.A., Liyanapathiranage, P., Iruegas-Bocardo, F., White, F.F., Goss, M.M. & Jones, J.B. (2020). *Xanthomonas* diversity, virulence and plant-pathogen interactions. *Nature Reviews Microbiology, 18*, 415-427.

Ullah, I., Sjöstrand, J., Andersson, P. & Sennblad, B. (2015). Integrating sequence evolution into probabilistic orthology analysis. *Systematic Biology, 64*(6), 969-982. DOI: 10.1093/sysbio/syv044

Vauterin, L., Hoste, B., Kersters, K. & Swings, J. (1995). Reclassification of *Xanthomonas*. *International Journal of Systematic and Evolutionary Microbiology, 45*(3), 472-489. DOI: 10.1099/00207713-45-3-472

Wang, Y., Coleman-Derr, D., Chen, G. & Gu, Y.Q. (2015). OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research, 43*, W78-W84. DOI: 10.1093/nar/gkv487

Wu, H.J., Wang, A.H.J. & Jennings, M.P. (2008). Discovery of virulence factors of pathogenic bacteria. *Current Opinion in Chemical Biology, 12*(1), 93-101. DOI: 10.1016/j.cbpa.2008.01.023