Dennis C. Gross · Ann Lichens-Park
Chittaranjan Kole *Editors*

Genomics of Plant-Associated Bacteria

Springer

Genomics of Plant-Associated Bacteria

Dennis C. Gross · Ann Lichens-Park
Chittaranjan Kole
Editors

Genomics of Plant-Associated Bacteria

Springer

Editors
Dennis C. Gross
Department of Plant Pathology
and Microbiology
Texas A&M University
College Station, TX
USA

Ann Lichens-Park
United States Department of Agriculture
National Institute of Food
and Agriculture
Washington, DC
USA

Chittaranjan Kole
Bidhan Chandra Krishi Viswavidyalaya
Mohanpur, West Bengal
India

ISBN 978-3-642-55377-6 ISBN 978-3-642-55378-3 (eBook)
DOI 10.1007/978-3-642-55378-3
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014940933

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The first genome of a plant pathogen (*Xylella fastidiosa* that causes citrus variegated chlorosis) was published in 2000. In that same year, the United States Department of Agriculture (USDA) first offered a competitive grants program to support sequencing of agriculturally relevant microorganisms. At that time, only a few microbes of agricultural significance were being sequenced. In 2001, the United States National Science Foundation (NSF) joined USDA, and the jointly-offered program was expanded to support the genomic sequencing of an even broader range of microorganisms. The partnership lasted until 2009, by which time sequencing costs had decreased dramatically and sequencing speed had increased enormously. USDA's support for microbial genomics shifted more towards functional analysis of genome sequences. During the 10 years that the genome sequencing program was offered, the genomes of a large number of agriculturally significant microorganisms were sequenced with support from the USDA, including bacteria, viruses, fungi, oomycetes, and even a nematode. The genomes of microbes relevant to basic science were sequenced with funding from the NSF. Three books are being published to describe the impact of some agriculturally relevant genomes and their analysis. In addition to this volume, two other volumes (edited by Dr. Ralph Dean, Dr. Ann Lichens-Park and Dr. Chittaranjan Kole) describe the genomic analysis of plant-associated fungi and oomycetes. These volumes are entitled "Genomics of Plant-Associated Fungi: Monocot Pathogens" and "Genomics of Plant-Associated Fungi and Oomycetes: Dicot Pathogens." This book describes how the availability of some agriculturally important plant-associated bacterial genomes, many of which were first sequenced with support from the USDA, have revolutionized our understanding of these bacteria and how they can be managed to improve the sustainability of agriculture worldwide.

Bacterial genome data is a launching pad (or scientific springboard) that, with appropriate functional analysis, can lead to greater knowledge of bacterial evolution and diversity, ecological adaptation, and environmental influences on gene expression and phenotype. Because

bacterial genomes are generally much smaller than the genomes of eukaryotic organisms, important insights about how the bacteria function can be obtained from appropriate functional analysis, often in less time than that needed for such analyses employed in higher eukaryotes. At the same time, studies aimed at understanding how bacteria interact with plants can lead to new insights about important plant processes and metabolic pathways, some of which may be exploited by bacterial pathogens to cause diseases. As is explained in the chapter on the genomics of *Xanthamonas oryzae*, the availability of xanthomonad genomic information has led to discoveries of new technologies such as transcription activator-like effector nucleases (TALENs) that facilitate targeted genome editing and will revolutionize approaches to genetic engineering in eukaryotes.

Each chapter in this book describes the genomic analysis of a particular bacterial genus, species or group of related bacteria about which the genome sequence and genomic analysis have led to significant new insights. Some chapters address bacterial pathogens that are readily tractable to genetic analysis and are, therefore, considered to be good model systems. *Pseudomonas syringae*, the subject of two chapters, is an important pathogen and a model system. Other chapters focus on bacterial pathogens that are less tractable genetically, but which are responsible for diseases that can result in devastating economic losses for growers. *Xanthamonas citri* causes citrus canker, a disease that can result in severe losses to citrus crops. One chapter focuses on *Pseudomonas flourescens*, which is a beneficial bacterial species that can help prevent diseases in plants. The very first plant pathogen, *Erwinia amylovora*, was first described in the 1800's and is the subject of one of the chapters. The fastidious phloem-limited bacteria, exemplified by the phytoplasmas and the liberibacters, were first described in recent years, and are each the subject of one chapter.

Several "themes" run through the chapters in this volume. These include the growing evidence for the importance of horizontal gene transfer or "nature's transgenes" in originating new bacterial strains and species. Advances in transcriptomic analysis are facilitating studies describing complex regulatory networks critical to expression of processes important in plant–microbe interactions. The technology is leading to identification of new bacterial factors or products that mediate communication with and establishment in the plant host. Genomic studies of plant-associated bacteria promise to lead to a better understanding of the natural microbial communities associated with plants (the phytobiome) and to innovative means of controlling diseases caused by plant pathogens.

We wish to express our thanks to the lead authors and co-authors of the chapters in this volume. They have done a marvelous job of

explaining the advances and significance of the new knowledge described in their chapters. We also wish to express our thanks to some special people who are current or former employees of USDA and NSF whose support has been critical to the microbial genomics program and to the existence of this volume. These people are Dr. Sonny Ramaswamy, Dr. Colien Hefferan, Ms. Betty Lou Gilliland, Ms. Erin Daly, Mr. Edward Nwaba, Dr. Deborah Sheely, Ms. Cynthia Montgomery, Dr. Michael Fitzner, Dr. Daniel Jones, Ms. Pushpa Kathir, Dr. Anna Palmisano, Dr. Mark Poth, Dr. Maryanna Henkart, and all of the USDA and NSF Program Officers and staff who worked with Dr. Lichens-Park while the Microbial Genome Sequencing Program was offered. Space limitations prevent us from describing each of the roles played by these individuals but they all have been significant and we are extremely grateful to all of them.

March 2014

Dr. Dennis C. Gross
Texas A&M University
College Station, TX, USA

Dr. Ann Lichens-Park
National Institute of Food and Agriculture
United States Department of Agriculture
Washington, DC, USA

Prof. Chittaranjan Kole
Bidhan Chandra Krishi Viswavidyalaya
West Bengal, India

Contents

Genomics of *Erwinia amylovora* and Related *Erwinia* Species Associated with Pome Fruit Trees

1

Youfu Zhao

1.1 Introduction

Erwinia amylovora, the causal agent of fire blight of apple, pear, quince, blackberry, raspberry, and other rosaceous plants, is of particular interest to plant bacteriologists, not only because it is the first bacterium demonstrated to cause disease in plants, but also because of its significant economic and political impact (Griffith et al. 2003; van der Zwet et al. 2012; Vanneste 2000). Recently, the pathogen was voted as one of the top 10 plant pathogenic bacteria in molecular plant pathology (Mansfield et al. 2012). In the last decade, several new pathogenic *Erwinia* species associated with pome fruit trees have been described, including *Erwinia pyrifoliae*, *Erwinia piriflorinigrans*, and *Erwinia uzenensis* (Kim et al. 1999; Matsuura et al. 2012; Lopez et al. 2011). Additionally, *Erwinia billingiae* and *Erwinia tasmaniensis* are two non-pathogenic *Erwinia* species associated with pome fruit trees (Geider et al. 2006; Mergaert et al. 1999). *E. amylovora* is the type species of the genus (Lelliott and Dickey 1984) and remains the most studied species of this genus. Less information is available on other *Erwinia* species associated with pome fruit trees and their genetic relationship to *E. amylovora* (Palacio-Bielsa et al. 2012). Resolution of the genetic compositions of these microorganisms has, therefore, dramatically increased our knowledge base of *E. amylovora* and its relatives (Zhao and Qi 2011). Complete and draft genome sequences for more than a dozen strains, belonging to five *Erwinia* species, including *E. amylovora*, *E. pyrifoliae*, *E. piriflorinigrans*, *E. tasmaniensis*, and *E. billingiae*, have been published (Kube et al. 2008, 2010; Mann et al. 2013; Park et al. 2011; Powney et al. 2011; Sebaihia et al. 2010; Smits et al. 2010a, b, 2013). These genome sequences provide almost complete genetic information about *E. amylovora* and other closely related species. In this chapter, we present (1) an updated review of *E. amylovora* and related species from genome sequencing efforts; (2) summarize the general characteristics of the pathogen, the disease it causes, and its genome; and (3) highlight current genome-enabled understanding of *E. amylovora* pathogenesis, including comparative genomic analyses and evolution, as well as genetic and functional genomic studies. Future perspectives and research directions for this important pathogen are also discussed.

1.1.1 Fire Blight and Related Diseases Associated with Pome Fruit Trees

Fire blight has been known as one of the most important plant bacterial diseases worldwide and is a devastating necrotic disease affecting apples, pears, and other rosaceous plants (Norelli et al.

Y. Zhao (✉)
University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA
e-mail: zhao888@illinois.edu

D. C. Gross et al. (eds.), *Genomics of Plant-Associated Bacteria*,
DOI: 10.1007/978-3-642-55378-3_1,

2003). Fire blight was first observed on rosaceous plants in 1780 and was considered the first recognized plant bacterial disease back in 1884 (van der Zwet and Keil 1979). Several bacterial diseases with symptoms resembling those of fire blight have been described for pome fruit trees in recent years. Bacterial shoot blight of pear, first described as a fire blight-like disorder, was observed on Asian pear (*Pyrus ussuriensis* cv. Mishirazu) and European pear (*Pyrus communis* L.) trees on Hokkaido island in Japan in the 1970s (Tanii et al. 1981), and the disease was considered eradicated in 1999 (Mizuno et al. 2000, 2010). A similar disease was described on Asian pear (*Pyrus pyrifolia* cvs. Shingo and Mansamgil) in Korea in the late 1990s (Kim et al. 1999). In 2007, bacterial black shoot disease of European pear cv. La France was found in Kaminoyama City, Yamagata Prefecture, Japan (Mizuno et al. 2010). At the same time, a necrotic pear blossom disease of European pear cvs. Ercolini (Coscia) and Tendral was observed in Valencia, Spain (Rosello et al. 2006). Due to their restricted distribution and limited access to disease samples for the fire blight research community, there is not much information about these new diseases beyond their original location. It is expected that these diseases may have a wider distribution than has been reported.

1.1.1.1 Fire Blight Disease and Its Current Distribution

Historically, a severe outbreak of fire blight in the summer of 1880 in Illinois led to the discovery of the first bacterial disease of plants by Thomas J. Burrill, a professor and former president of the University of Illinois at Champaign-Urbana (Griffith et al. 2003). The disease is indigenous to North America and was first observed on pear and quince (*Cydonia oblonga* Mill) in the lower Hudson Valley of New York in 1780 (Griffith et al. 2003). Since 1862, the disease has been widespread and destructive in commercial apple (*Malus sylvestris* Mill) and pear orchards in the northeastern United States. In 1882, the disease spread from New York to Illinois and to California in 1887 (Kado 2000; Pierce 1902). Another milestone in fire blight research occurred in 1891, when insects were demonstrated to be attracted to exudates from infected shoots in the spring and subsequently transmitted the disease to other trees (Kado 2000). This is also the first (bacterial) plant pathogen for which an insect vector was demonstrated.

Since its discovery, fire blight has spread to more than 50 countries (van der Zwet et al. 2012). Fire blight was first observed in Ontario, Canada, in 1904; in British Columbia in 1911; and in Mexico in 1921. Since 1924, the disease has spread to all pear- and apple-growing areas of Canada and Mexico. The reported long-distance spread of fire blight from America was to New Zealand in 1919; to the United Kingdom and Northern Europe in the late 1950s (Crosse et al. 1958); and to the Mediterranean Region and Northern Africa in the 1960s. Fire blight is now widespread across Europe, Northern Africa, and the Middle East including Iran, threatening the native origin of apple germplasm resources in central Asia (van der Zwet et al. 2012). Interestingly, fire blight has not been reported in Australia, China, South Africa, and any country in South America. However, fire blight-related diseases associated with pome fruit trees were reported in Japan and South Korea (Kim et al. 1999; Mizuno et al. 2010).

1.1.1.2 Economic Losses and Costs of Fire Blight Disease Management

Since its discovery, fire blight has been considered the most destructive disease and a limiting factor for apple and pear production (Palacio-Bielsa et al. 2012). Fire blight not only can greatly reduce crop yield and marketability in the current season by infecting blossoms and killing of fruit spurs, but also cause the loss of entire trees and orchards. The severity of fire blight outbreaks in California, and the Midwest prevents the commercial production of pears in these regions (Eastgate 2000; Pierce 1902). The increased replanting cost of high-density blocks (ca. $20,000 per ha) and losses due to international trade regulations on fire blight has resulted

in significant financial losses for many growers. Therefore, fire blight poses multifarious threats to the global pome fruit industry.

It is extremely difficult to accurately estimate the economic losses due to fire blight to an individual grower, a region, and on a global scale. However, some of the most traumatic fire blight epidemics recorded in recent years provide a glimpse into how fire blight can cause great economic losses. In 1998, losses were estimated to be in excess of $68 million in Washington and northern Oregon (van der Zwet et al. 2012). In 2000, the most widespread epidemic of fire blight occurred in southern Michigan and losses were estimated to be $42 million, including removal of 300,000 trees and replacement of 1,550 acres of young orchards (Longstroth 2000).

Fire blight has also become a great threat to the world's pome fruit industry. Losses were estimated at 10 million dollars (NZ) in the Hawke's Bay region of New Zealand (Vanneste 2000). In the 1990s, more than a half million trees were destroyed in Italy alone (Vanneste 2000). In 2007, a severe outbreak of fire blight occurred in Switzerland. The Swiss government reported a loss of $27.5 million dollars and 10 % of their apple acreage (Ashton 2008). Fire blight also affects the ornamental nursery business. The most popular ornamental plant in Europe in the 1970s was cotoneaster, a major host of fire blight. In 1975, more than 2 million cotoneasters were destroyed in nurseries and garden centers in the Netherlands (van der Zwet et al. 2012).

Since the 1970s, spray applications of streptomycin have been the most effective means of controlling blossom blight of apples and pears. However, the occurrence of streptomycin resistance in the USA, Canada, and elsewhere has rendered this antibiotic ineffective (Chiou and Jones 1995; Coyier and Covey 1975; McManus et al. 2002). Moreover, the use of streptomycin for control and the progressive accumulation of resistant strains have been estimated to bring additional losses of more than 100 million dollars (US) per year in the USA (Norelli et al. 2003). Furthermore, fire blight also results in great economic losses due to stringent quarantine and international trade regulations. For example, Australia, free of fire blight, reported spending $40 million Australian dollars for the eradication, diagnostics, loss of sales, and exports due to an unconfirmed presence of fire blight (Rodoni et al. 2006). Moreover, strict regulatory measures against *E. amylovora* are still imposed by many European countries, such as Switzerland, Belgium, and Germany (Deckers 1996; Duffy et al. 2005; EPPO 1992). In Germany, expenses for eradication, inspection, fruit losses, and chemical treatment of fire blight totaled $110 million from 1972 to 1990 (van der Zwet et al. 2012).

1.1.1.3 Disease Symptoms and Life Cycle

The disease name "fire blight" was first coined by William Coxe in 1817, when he described a disorder that "in a few hours, turn the leaves suddenly brown, as if they had passed through a hot flame" (Griffith et al. 2003). Now, we can easily recognize that this is the typical symptoms of fire blight, i.e., the appearance of blackening tissue as though they had been scorched by fire (Schroth et al. 1974). In nature, *E. amylovora* is capable of infecting blossoms, fruits, vegetative shoots, woody tissues, and rootstock crowns, leading to blossom blight, fruit blight, shoot blight, twig and trunk blight, and rootstock blight symptoms with frequent creamy ooze production (Eastgate 2000). Usually, blossom blight is the first symptom of the disease in early spring followed by infection of young fruits (Thomson 1986). Succulent shoots, twigs, and water sprouts are the next most susceptible parts of the tree. Infected young succulent shoots and twigs wither and turn brown and in most cases, the tip of the shoot bends in a characteristic fashion to form a "shepherd's crook." From the infected blossoms, shoots, or fruits, the disease spreads systemically through the spurs to larger twigs and branches to cause cankers. Disease then may spread into the scaffold limbs and the trunk, resulting in the death of the entire tree. The fire blight disease cycle has been extensively described by a number of authors

(Thomson 1986; van der Zwet and Keil 1979; van der Zwet et al. 2012).

Bacterial shoot blight disease, bacterial black shoot disease, and necrotic pear blossom disease induce very similar symptoms to those of fire blight, but major differences exist (Palacio-Bielsa et al. 2012). The symptoms of bacterial shoot blight disease of Asian pear, caused by *E. pyrifoliae*, include black to brown stripes in the leaf midribs, dark brown leaf spots, and necrotic petioles on large parts of the trees (Kim et al. 1999; Rhim et al. 1999). However, the symptoms of black shoot disease of European pear trees, caused by *E. uzenensis* (Matsuura et al. 2012), were different from those of bacterial shoot blight disease (Rhim et al. 1999) and fire blight. Necrotic symptoms were only observed in young shoots, and the development of lesions stopped within 20 cm from the base of the shoots and did not affect the branches (Mizuno et al. 2000). Some typical symptoms of fire blight and bacterial shoot blight disease (i.e., blossom blight, fruitlet blight, and formation of a shepherd's crook), were not observed for black shoot disease (Matsuura et al. 2012). Furthermore, the symptoms of necrotic pear blossom disease, caused by *E. piriflorinigrans*, were also different from typical fire blight disease (Rosello et al. 2006; Lopez et al. 2011). Only pear blossoms showed necrotic symptoms, but not pear shoots or fruitlets, apple trees, or other inoculated *Rosaceae* species (Rosello et al. 2006).

1.1.2 The Pathogen *E. amylovora* and Related *Erwinia* Species

1.1.2.1 General Characteristics

E. amylovora and related *Erwinia* species all belong to the family *Enterobacteriaceae* and the genus *Erwinia*. *E. amylovora* was initially described as *Micrococcus amylovorus* by Professor Burrill in 1883 and then *Bacillus amylovorus* (Burrill) by Trevisan in 1889, under the erroneous assumption that the pathogen hydrolyzes starch. In the early 1900s, it was renamed as *E. amylovora* (Burrill) by Winslow et al. (Lelliott and Dickey 1984). Though *E. amylovora* and related *Erwinia* species share many basic characteristics, such as Gram negative; rod shaped; motile with peritrichous flagella, facultative anaerobic growth, oxidase negative, catalase positive; and acid production from glucose, fructose, and galactose (Palacio-Bielsa et al. 2012), they differ in some phenotypic traits, host range, and virulence factors (Kim et al. 1999; Lopez et al. 2011; Mizuno et al. 2010; Rhim et al. 1999; Rosello et al. 2006; Shrestha et al. 2003).

Host Range of *E. amylovora* and Related *Erwinia* Species

One major difference between *E. amylovora* and related *Erwinia* species is their abilities to cause disease on different host plants (host range). *E. amylovora* has a wide host range within the family of *Rosaceae*. However, other pathogenic *Erwinia* species associated with pome fruit trees are host-specific, only infecting certain varieties of Asian or European pear trees, and some with tissue specificity. For example, *E. piriflorinigrans* only infects blossoms of European pear trees (Rosello et al. 2006), whereas *E. uzenensis* causes disease on young shoots of European pear trees with limited disease progress after infection (Mizuno et al. 2010). The host range of *E. pyrifoliae* may be broader, as disease symptoms are observed on several varieties of Asian pears and after inoculation, on several commercial European pear cultivars and apple (*Malus domestica*) (Kim et al. 2001; Mizuno et al. 2010). Furthermore, some *E. amylovora* strains isolated from *Rubus* plants within the subfamily *Rosoideae* are also host-specific, which can only infect *Rubus* plants. Interestingly, natural *E. amylovora* strains isolated from *Maloideae* (*Spiraeoideae*) with a wide host range show differential virulence on different apple cultivars (Lee et al. 2010; Wang et al. 2010a). These observations and early genetic studies suggest that *E. amylovora* strains may further be divided into different species or subspecies with distinct host ranges, i.e., strains isolated from *Maloideae* and *Rosoideae*.

The host range of *E. amylovora* includes more than 180 species from 39 genera in the family of *Rosaceae* (van der Zwet et al. 2012) and all four subfamilies (old classification system): *Maloideae* (syn. *Pomoideae*), *Rosoideae*, *Amygdaloideae* (syn. *Prunoideae*), and *Spiraeoideae*. Of the 39 genera, eight are fruit crops: *Malus*, *Pyrus*, *Cydonia, Eriobotrya, Fragaria, Mespilus, Prunus,* and *Rubus* (van der Zwet 1995). The remaining genera are mostly ornamental plants and trees, including those most susceptible and economically important hosts: *Cotoneaster*, *Crataegus*, *Pyracantha*, and *Sorbus*. Fire blight has been described in raspberry (*Rubus idaeus*) (Starr et al. 1951), in *Rosa rugosa* in Germany (Vanneste et al. 2002), and in chokeberry and strawberry in Bulgaria (Bobev et al. 2007). The complete list of host plants for fire blight could be found in van der Zwet et al. (2012) and van der Zwet and Keil (1979).

Taxonomic Position and Phylogenetic Relationships

Taxonomically, the genus *Erwinia* belongs to the *γ-Proteobacteria* (Order *Enterobacteriales*, Family *Enterobacteriaceae*) (Starr and Chatterjee 1972). Based on sequence analysis of the 16S rRNA genes, *E. amylovora* and related species associated with pome fruit trees are closely related to other enterobacteria such as *Escherichia coli*, *Salmonella enterica*, and *Yersinia pestis*. A phylogenetic tree based on 16S rRNA gene sequences showed that all strains of *E. amylovora* formed a separate clade within the genus *Erwinia* (Matsuuza et al. 2012). Housekeeping genes, such as *atpD*, *gyrB*, *infB,* and *rpoB* are commonly used for the phylogenetic analysis of *Enterobacteriaceae* (Sarkar and Guttman 2004). A phylogenetic tree reflecting their evolutionary relationship from concatenated sequences of four housekeeping proteins (AcnB, GltA, GyrB, and RpoD) is presented in Fig. 1.1 (Zhao and Qi 2011). The topologies of the phylogenetic trees based on 16S rRNA gene and housekeeping genes are very similar. Furthermore, the topology is identical to a phylogenetic tree generated using core genomes of the

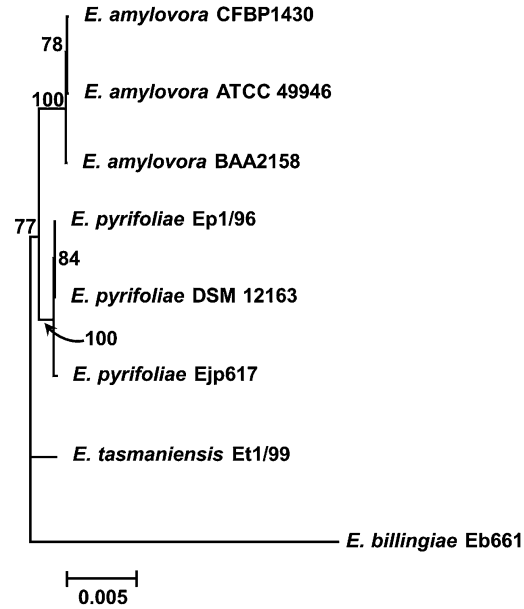

Fig. 1.1 Evolutionary relationship of *Erwinia amylovora* and related *Erwinia* species from Zhao and Qi (2011). The phylogenetic tree was constructed from concatenated sequences (2222 aa) of four housekeeping proteins (AcnB, GltA, GyrB and RpoD) using the neighbor-joining (NJ) method. Bootstrap scores greater than 60 are given at each node. The scale bar represents 0.005 amino acids substitutions per site. *E. billingiae* strain Eb661 was used as an outgroup

sequenced *E. amylovora* strains (Mann et al. 2013). In addition, *E. uzenensis* and *E. piriflorinigrans* cluster more closely to *E. pyrifoliae* strains than to *E. amylovora* strains (Matsuuza et al. 2012). Within the species of *E. amylovora*, phylogenetic analyses of the core genomes of *E. amylovora* strains indicate that the *Spiraeoideae*-infecting strains exhibit much less diversity than the *Rubus*-infecting strains (Mann et al. 2013; Rezzonico et al. 2012). This grouping is consistent with previous studies using rep-PCR, carbon utilization, and phylogeny based on *rpoB* (Mann et al. 2013; Rezzonico et al. 2012).

1.1.2.2 Virulence Factors and Regulation

E. amylovora is a highly virulent necrogenic vascular pathogen. Extensive genetic studies over the past several decades have explored the

molecular mechanism of *E. amylovora* pathogenesis (Khan et al. 2012; Zhao and Qi 2011). One of the highlights is the identification of an essential virulence system, namely the hypersensitive response and pathogenicity (*hrp*)-type III secretion system (T3SS), which is clustered on an ostensible pathogenicity island 1(PAI1) (Oh and Beer 2005; Oh et al. 2005). The current understanding is that two virulence factors are strictly required for *E. amylovora* to cause disease, i.e., the T3SS and its effectors and the acidic exopolysaccharide (EPS) amylovoran (Zhao and Qi 2011; Zhao et al. 2009a, b). Other virulence determinants found in *E. amylovora* that contribute to virulence and plant colonization include levan, protease, the iron-scavenging siderophore desferrioxamine, and genes involved in sorbitol and sucrose metabolism (Aldridge et al. 1997b; Bogs and Geider 2000; Dellagi et al. 1999; Du and Geider 2002; Smits and Duffy 2011).

Type III Secretion and Effectors

The discovery of the *hrp* gene cluster, which encodes a T3SS common to Gram-negative bacterial pathogens, is a breakthrough in understanding the molecular mechanism of bacterial pathogenesis and is a landmark in modern molecular plant pathology (Lindgren et al. 1986). The most noticeable phenotype of *hrp* mutants is their lost pathogenicity on susceptible host plants, and inability to elicit the hypersensitive response (HR) on resistant cultivars of host plants and non-host plants (Barny et al. 1990; Lindgren et al. 1986; Steinberger and Beer 1988). In *E. amylovora*, the PAI1 *hrp* gene cluster is located on an approximately 60-kb genomic region (Bogdanove et al. 1996; Kim and Beer 2001). The *hrp* gene cluster can be divided into three subregions: the *hrp/hrc* region, the Hrc effector and elicitors (HEE) region, and the Hrp-associated enzymes (HAE) region (Oh and Beer 2005; Oh et al. 2005). The *hrp/hrc* region contains 25 genes, including four regulatory genes (*hrpL*, *hrpS* and *hrpXY*) and genes encoding structural components of T3SS. Among them, nine *hrc* genes constitute the core structural components of the T3SS, including HrpA, a pilin protein (Jin et al. 2001; Kim et al. 1997). The seven gene HEE regions encode two harpin proteins (HrpN and HrpW) and several effectors and chaperones, including disease-specific protein DspE/A and its chaperone DspF/B (Bogdanove et al. 1998a, b; Gaudriault et al. 1997). The HAE region contains five genes, including three *hrp*-associated systemic virulence genes (*hsvABC*) and *hrpK*, which encodes a putative translocator (Oh et al. 2005). Therefore, T3SS proteins can be categorized into three groups, regulatory proteins (e.g., HrpL sigma factor), secretion apparatus proteins (e.g., HrpA pilin), and extracellularly secreted effector proteins (e.g., HrpN harpin and DspE).

E. amylovora has been developed as a model pathogen for studying plant–microbe interactions because the first cell-free elicitor (HrpN, harpin) was identified in 1992 (Wei et al. 1992). The T3SS of *E. amylovora* secretes at least 15 virulence-associated proteins, including HrpA, HrpN, HrpW, HrpJ, HrpK, HopAK1 (Eop2), DspE, HopC1, HopX1 (Eop3); AvrRpt2 (Eop4), and Eop1(EopB, OrfB) (Bogdanove et al. 1998a, b; Gaudriault et al. 1997, 1998; Nissinen et al. 2007; Zhao et al. 2005, 2006). Among them, harpins (HrpN and HrpW) are glycine-rich hydrophilic proteins and elicit an HR when infiltrated into intercellular spaces of some plants (Wei et al. 1992). HrpJ, HrpK, and HopAK1 (Eop2, Eam_2780) are putative translocators for delivery of effector proteins or Hop proteins (*H*rp *o*uter *p*roteins). Together with HrpN, they may form the translocon apparatus.

Many studies including genome sequencing have reached the conclusion that only five effector or *hop* genes [*eop1*, *eop3* (*hopX1*), *eop4* (*avrRpt2*), *dspA/E*, and *hopPtoC* (*hopC1*)] and several chaperones (*dspF/B, esc1* (*orfC*), and *esc3*) exist in the genome of *E. amylovora* (Nissinen et al. 2007; Zhao et al. 2005, 2006). Among them, *eop3* (*Eam_2190*), *eop4* (*Eam_0423*), *eop2*, and *hopC1* (*Eam_2679*) are located outside of the T3SS PAI1 in the genome. *DspA/E*, *avrRpt2*, and *hopC1* have been demonstrated to be induced in immature pear fruit,

indicating that they may play a major role in virulence (Zhao et al. 2005, 2006). Eop1 and Eop3 are AvrRxv/YopJ and HopX family proteins, respectively. Eop3 (HopX1) may function as an avirulence gene in apple (Bocsanczy et al. 2012), whereas Eop1 may function as a factor in limiting host range (Asselin et al. 2011).

Exopolysaccharide Amylovoran

E. amylovora produces two types of EPSs, the heteroexopolysaccharide amylovoran and the homoexopolysaccharide levan (Bellemann and Geider 1992; Geier and Geider 1993). Levan is a homopolymer of fructose residues and is synthesized by the enzyme levansucrase. Mutants deficient in levan production are reduced in their virulence (Geier and Geider 1993; Gross et al. 1992). However, strains deficient in levan production have also been found in nature and are virulent (Bereswill et al. 1997). Amylovoran may have multiple functions because mutants deficient in amylovoran biosynthesis are nonpathogenic (Bellemann et al. 1994; Bernhard et al. 1993). Furthermore, the ability of individual *E. amylovora* strains to produce amylovoran is positively correlated with the degree of virulence (Wang et al. 2010a). More recent studies also revealed the indispensable functions of amylovoran in bacterial biofilm formation in plant xylem and for survival under stress conditions (Koczan et al. 2009; Ordax et al. 2010).

Amylovoran was first isolated from bacterial ooze (Bennet and Billing 1980). It is a complex, high molecular weight (50–150 mDa), acidic capsular EPS, consisting of galactose, glucose, and pyruvate residues (Nimtz et al. 1996). Amylovoran biosynthetic genes are located within a 12-gene amylovoran biosynthetic (*ams*) operon, from *amsA* to *amsL*, with *amsG* as the first gene in the operon (Aldridge et al. 1997a; Bernhard et al. 1993; Bugert and Geider 1995). Another two genes, *galF* and *galE*, which are located on the right adjacent to the *ams* cluster, are involved in amylovoran precursor formation. Studies have proposed that products of the *amsGBCDEJK* genes play roles in glycosyl transfer for the repeating unit (Langlotz et al. 2011). Until recently, new evidence has shown that ams-II (*amsG2*) and ams-III (*amsO-amsL2*) may also play a role in side-chain modification (Wang et al. 2012c).

Regulatory Systems

In prokaryotes, gene expression is regulated primarily at the level of transcription initiation. In *E. amylovora*, transcription of the *hrp*-T3SS genes is activated by the master regulator HrpL, a member of the ECF subfamily of sigma factors (Wei and Beer 1995). HrpL binds to a consensus sequence known as the *hrp* box (GGAACC-N_{16}-CCACNNA) in *hrp* gene promoters. Most T3SS and effector genes are subject to direct HrpL regulation (McNally et al. 2012, Nissinen et al. 2007). A hidden Markov model has identified about 30 *hrp* promoters in the genome of *E. amylovora* strain Ea273, which contain the *hrp* box recognized by HrpL (McNally et al. 2012; Bocsanczy et al. 2012).

The expression of *hrpL* is believed to be activated by both HrpS and a two-component regulatory system HrpX (sensor) and HrpY (response regulator) (Wei et al. 2000). Further domain structure analysis indicated that HrpX contains two PAS domains (initially found in *P*ER, *A*RNT, and *S*IM proteins) within the N-terminal sensor region, suggesting that HrpX is a soluble and cytoplasmic protein that may sense intracellular signals. Recent studies also found that *hrpXY* mutants remain virulent (Zhao et al. 2009b), which is different from a previous report that analyzed Tn*5*-insertional *hrpXY* mutants (Wei et al. 2000). An early report suggested that HrpS, a member of the NtrC family of σ^{54} enhancer-binding proteins, only partially controls *hrpL* expression (Wei and Beer 1995). However, recent findings indicated that HrpS, YhbH, and alternative sigma factor RpoN (σ^{54}) are absolutely required for *hrpL* expression (Ancona et al. 2014; Zhao et al. 2009b) and the *hrpL* gene also contains a σ^{54} consensus sequence in its promoter region. Furthermore, global regulators such as GrrSA and EnvZ/OmpR two-component systems may also be involved in regulating *hrp* gene expression (Li et al. 2014).

Several key regulators of amylovoran biosynthesis have been characterized earlier, including RcsA (Bernhard et al. 1990; Coleman et al. 1990) and RcsB (Bereswill and Geider 1997) as well as interactions between RcsA and RcsB (Kelm et al. 1997; Wehland and Bernhard 2000; Wehland et al. 1999). Recently, the RcsCDB system has been demonstrated to be essential for virulence (Wang et al. 2009, 2011b; Zhao et al. 2009b). The Rcs phosphorelay system is a unique enterobacterial-specific two-component system, and phosphorylated RcsB could form RcsB-RcsB homodimers or interact with RcsA to form RcsAB heterodimers, which then bind to an "RcsAB box" to regulate gene expression, including the promoter of the *ams* operon involved in amylovoran biosynthesis (Pristovsek et al. 2003; Wehland et al. 1999, Wehland and Bernhard 2000). A Hidden Markov model identified about 60 genes in the genome of *E. amylovora* strain Ea273 (ATCC49946), which contains the RcsAB box, and half of these genes were directly regulated by RcsBC (Wang et al. 2012a). Genome-wide screening of two-component-system mutants identified four groups of mutants that exhibited varying levels of amylovoran production in vitro, indicating that two-component systems in *E. amylovora* play a major role in regulating amylovoran production, and may form a regulatory network to govern the production of amylovoran (Wang et al. 2011c; Zhao et al. 2009b).

Additional novel and global regulatory genes for amylovoran biosynthesis have also been identified through genetic screening. These included Lon protease, global regulator H-NS, RcsF, DjlA, and AmyR (YbjN) (Eastgate et al. 1995; Hildebrand et al. 2006; Wang et al. 2011c, 2012b). Both RcsF and DjlA are activators of RcsC, whereas H-NS binds to the promoter of *rcsA* and suppresses *rcsA* gene expression. Furthermore, the RcsA protein is subject to Lon-dependent degradation, which is a heat-shock protein. In addition, AmyR, an *amy*lovoran *r*epressor, is an enterobacterial-specific orphan protein and was recently characterized as a novel negative regulator of EPS production in both *E. coli* and *E. amylovora* (Wang et al. 2011a, 2012b).

Based on current knowledge, a simple model for *E. amylovora* virulence gene expression centered on T3SS and amylovoran production is presented in Fig. 1.2. Upon initiating plant infection, *E. amylovora* senses the unknown host/environmental signals. This process activates a sigma factor cascade that regulates T3SS gene expression and at the same time, several two-component signal transduction systems that regulate both T3SS and amylovoran production (Li et al. 2014; Zhao et al. 2009b). In the sigma factor cascade, σ^{54} enhancer-binding protein HrpS forms a hexamer and binds to the upstream DNA activator sequences (UAS) of the *hrpL* promoter. Meanwhile, RpoN and a core RNA polymerase (RNAP) forms a σ^{54}-RNAP complex that binds to the σ^{54} promoter of the *hrpL* gene, but remains transcriptionally silent. With the assistance of integration host factor (IHFα/β, not shown), HrpS contacts the σ^{54}-RNAP-promoter complex via the consensus GAFTGA motif and by DNA looping, with the energy provided by ATP hydrolysis of HrpS AAA+ -domain. This triggers the opening of the σ^{54}-RNAP-promoter complex and DNA melting. This process also requires a ribosome-associated protein YhbH with an unknown mechanism (Ancona et al. 2014). HrpL/RNAP complex then recognizes the "*hrp* box" at the promoter regions of HrpL-dependent operons or genes and regulates *hrp* gene expression (Wei et al. 2010). On the other hand, phosphorylated RcsB dimer or RcsAB heterodimer binds directly to the *amsG* promoter in the *ams* operon and regulates amylovoran production (Wang et al. 2009). The GrrSA two-component system specifically regulates small regulatory RNA *rsmB*, which stabilizes RNA-binding protein RsmA (Ancona and Zhao 2013; Li et al. 2014). However, the molecular mechanism as how GrrSA along with EnvZ/OmpR system negatively regulate T3SS and amylovoran production is unknown. Furthermore, the identity of the signals and how the bacterium senses the signals remain unsolved mysteries.

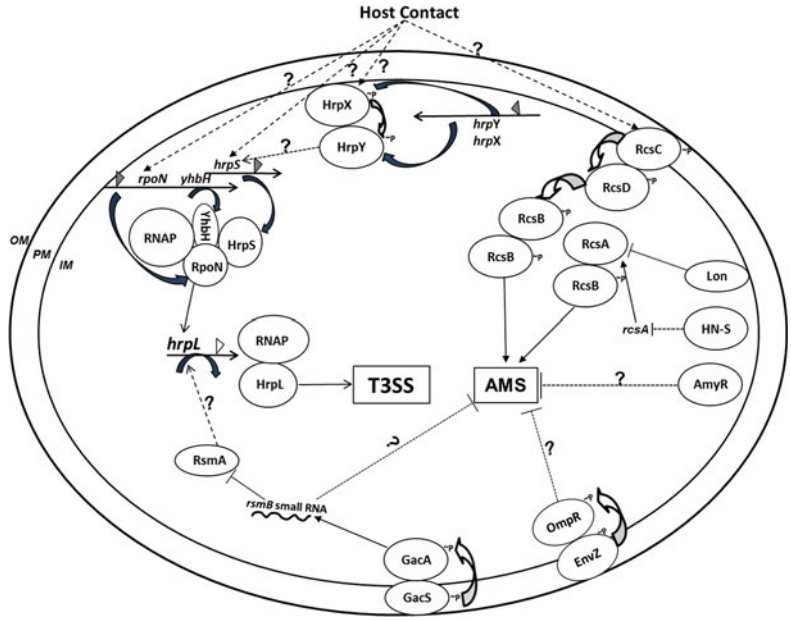

Fig. 1.2 A model for virulence gene regulation in *Erwinia amylovora*. AmyR orphan protein, ortholog of *E. coli* YbjN; AMS amylovoran biosynthesis, HN-S nucleoid-associated protein; HrpL; an ECF sigma factor, HrpS a σ^{54}-dependent enhancer-binding protein, HrpX/HrpY, OmpR/EnvZ, GacS/GacA, RcsABCD two-component regulatory systems, Lon protease, RpoN a sigma 54 factor, RNAP RNA polymerase, RsmA RNA-binding protein, *rsmB* a small regulatory RNA, T3SS type III secretion system, YhbH σ^{54} modulation protein (ribosome-associated protein). *OM* outer membrane, *PM* plasma membrane, *IM* inner membrane, *P* phosphorylation, *filled triangle*, σ^{70} promoter, *open triangle* σ^{54} promoter. Positive regulation is indicated by an *arrow* and down-regulation by T sign; question mark and *dash line* unknown mechanism; *Thick arrow* lines show gene or operons, *oval* and *circles* indicate proteins

1.1.2.3 Tools for Molecular Genetics Studies

Traditionally, genetic studies for plant pathogenic bacteria utilize mobile genetic elements such as Tn*5*, Mu, and Tn*10* to generate mutant libraries (Vanneste et al. 1990). However, because of the randomness, we may not obtain a mutant for a specific gene of interest. Alternatively, standard homologous DNA recombination systems using suicide vectors could also be used to generate mutants (Zhao et al. 2005); however, standard recombination techniques require extensive and time-consuming in vitro cloning steps (Zhao et al. 2005). Recently, we have successfully adopted a PCR-based one-step inactivation of chromosomal genes and also referred to as the Red-cloning technique, to generate mutants in *E. amylovora* (Datsenko and Wanner 2000; Zhao et al. 2006, 2009a, b). We found that this technique is not only easy and fast (capable of mutant generation within a week without any cloning steps), but can also generate deletion mutants for a single gene, an operon or a genomic island (Zhao et al. 2009a, b). These mutants provide basic tools to study gene function and signal transduction in *E. amylovora*.

1.2 Genomics

1.2.1 Genome Sequencing of *E. amylovora* and Related *Erwinia* Species Associated with Pome Fruit Trees

In recent years, complete genomes for five species from the genus *Erwinia*, including two *E. amylovora* strains, three *E. pyrifoliae* strains, one *E. tasmaniensis* strain, one *E. billingiae* strain, and one strain of *E. piriflorinigrans* from

Table 1.1 Overview of complete genome sequencing of *Erwinia amylovora* and related *Erwinia* species associated with pome fruit trees

Strains	Origin	Size[a] (Mb)	G + C content	Total proteins	Plasmid #s	Host	Accession #s
E. amylovora CFBP1430	France 1972	3.81	53.6	3,706	1	*Crataegus*	FN434113-114
E. amylovora ATCC 49946	New York 1973	3.81	53.6	3,565 (3,712)[b]	2	Apple	FN666575-577
E. pyrifoliae Ep1/96	South Korea 1996	4.03	53.4	3,697	4	Asian pear	FP236842 FP928999 FP236827-29
E. pyrifoliae DSM 12163 (Ep16/96)	South Korea 1996	4.03	53.4	4,038	4	Asian pear	FN392235-39
E. pyrifoliae Ejp617	Japan	3.91	53.6	3,672	5	Asian pear	CP002124-29
E. tasmaniensis Et1/99	Australia 1999	3.88	53.7	3,622	5	Apple flower	CU468128, 30-33, 35
E. billingiae Eb661	UK	5.10	55.2	4,917	2	Tree	FP236826, 30, 43
E. piriflorinigrans CFBP 5888	Spain 2000	3.97	49.8	3,857	1	Pear	CAHS0100001-25; HE792893

Data adapted from Kube et al. (2008, 2010), Mann et al. (2012, 2013), Park et al. (2011), Powney et al. (2011), Sebaihia et al. (2010), Smits et al. (2010a, b, 2013)

[a] Chromosomal size; plasmids are not included

[b] See Mann et al. 2013

Spain, have been sequenced (Table 1.1). In addition, draft genome sequences have been obtained for an additional 10 *E. amylovora* strains, including three strains isolated from *Rubus* spp. (Table 1.2). These genome sequences provide abundant scientific information about the genetic composition of these species and facilitate functional and comparative genomic studies to determine how genomes of closely related *Erwinia* species have evolved (Mann et al. 2013; Zhao and Qi 2011).

1.2.1.1 General Features and Distinguishing Characteristics

The first genome sequence of *E. amylovora* strain Ea273 (ATCC49946), funded by the United States Department of Agriculture, was started in the early 2000s and published in 2010 (Sebaihia et al. 2010). At the same time, the complete genome sequence of *E. amylovora* strain CFBP1430 was also published (Smits et al. 2010b). Draft genome sequences were later obtained and published for 10 additional *E. amylovora* strains, including three isolated from host-specific *Rubus* spp. (Mann et al. 2013; Powney et al. 2011). These sequenced strains represent various host plants, geographical origins, and temporal distributions to exemplify distinct bacterial populations of *E. amylovora* (Tables 1.1 and 1.2). Other general information, including genome size, GC content, total proteins, and plasmid content, is listed in Tables 1.1 and 1.2.

The genomes of *E. amylovora* and related *Erwinia* species associated with pome fruit trees range from 3.8 to 5.1 Mbp, with *E. amylovora* containing the smallest genome compared to other pathogenic enterobacteria sequenced so far (up to 5.5 Mbp) (Toth et al. 2006). The genome of *E. billingiae* is larger than that of pathogenic *Erwinia* species and the non-pathogenic *E. tasmaniensis*. A comparison of genomes of *E. amylovora* strains CFBP1430 and ATCC49946 shows that the two genomes share more than 99.9 % identity at the nucleotide level, indicating that *E. amylovora* is a relatively homogeneous

Table 1.2 Overview of draft genome sequencing of *Erwinia amylovora*

Strains	Origin	Size[a] (Mb)	G + C content	Total proteins	Plasmid #s	Host	Accession #s
E. amylovora CFBP1232	UK 1959	3.77	53.6	3,780	1	Pear	CAPB01000001-42; HF560650
E. amylovora 01SFR-BO	Italy 1991	3.77	53.6	3,744	1	*Sorbus*	CAPA01000001-11,HF560647
E. amylovora ACW 56400	Switzerland 2007	3.77	53.6	3,758	2	Pear	AFHN01000001-22, CP002951, AFHN01000023
E. amylovora UPN527	Spain 1997	3.77	53.6	3,746	0	Apple	CAPC01000001-18
E. amylovora CFBP2585 (Ea495)	Ireland 1986	3.77	53.6	3,734	2	*Sorbus*	CAOZ01000001-12, HF560645-46
E. amylovora EA266 (E4001A)	Canada	3.76	53.6	3,804	1	Apple	CAOY01000001-38, HF560644
E. amylovora Ea356 (Ea1/79)	Germany 1979	3.76	53.6	3,744	1	Cotoneaster	CAOX01000001-14,HF560643
E. amylovora BAA2158 (IL5)	Illinois 1972	3.81	53.6	3,827	3	*Rubus*	FR719181 to FR719212
E. amylovora Ea644	MA, USA 2003	3.80	53.3	3,937	1	*Rubus*	CAPD01000001-40, HF560648
E. amylovora MR-1(Ea574)	Michigan	3.79	53.4	4,042	1	*Rubus*	CAPE01000001-29, HF560649

Data adapted from Kube et al. (2008, 2010), Mann et al. (2012, 2013), Park et al. (2011), Powney et al. (2011), Sebaihia et al. (2010), Smits et al. (2010a, b)

[a] Chromosome size; plasmids are not included

species (Smits et al. 2010b; Zhao and Qi 2011). The chromosomes of another seven *Spiraeoideae*-infecting strains are also highly homogeneous. Greater genetic diversity is observed between *Spiraeoideae*- and *Rubus*-infecting strains than among individual *Rubus*-infecting strains; this suggests that further delineation of *E. amylovora* species with different host specificity may be warranted (Mann et al. 2013; Zhao and Qi 2011). Similarly, the genomes of the two *E. pyrifoliae* strains from Korea (Ep1/96 and DSM12163) are almost identical (Kube et al. 2010; Park et al. 2011; Smits et al. 2010a; Thapa et al. 2013).

One interesting characteristic of the *E. amylovora* genome is that five of seven copies of the rRNA operon have a 99-bp insertion within helix 45 of the 23S rRNA gene, which represents an intervening sequence (IVS) as described for some species in the *Enterobacteriaceae* family. Although IVSs are absent in *E. coli*, *Pectobacterium*, and *Dickeya* spp. (McGhee et al. 2002; Pronk and Sanderson 2001), IVS sequences are present in all seven copies of the 23S rRNA gene of *E. pyrifoliae* DSM 12163T, but are absent in copies of the 23S rRNA gene of *E. tasmaniensis* Et1/99 (Kube et al. 2008; Smits et al. 2010a). The rRNA fragmentation pattern from the excision of IVSs is shown in Fig. 1.3 for *E. amylovora*.

1.2.1.2 Genome Structure and Rearrangements

Whole-genome structure comparisons of *E. amylovora*, *E. pyrifoliae*, *E. piriflorinigrans*, *E. billingiae* and *E. tasmaniensis* with complete genome sequences reveal many large-scale chromosomal re-organizations and inversions,

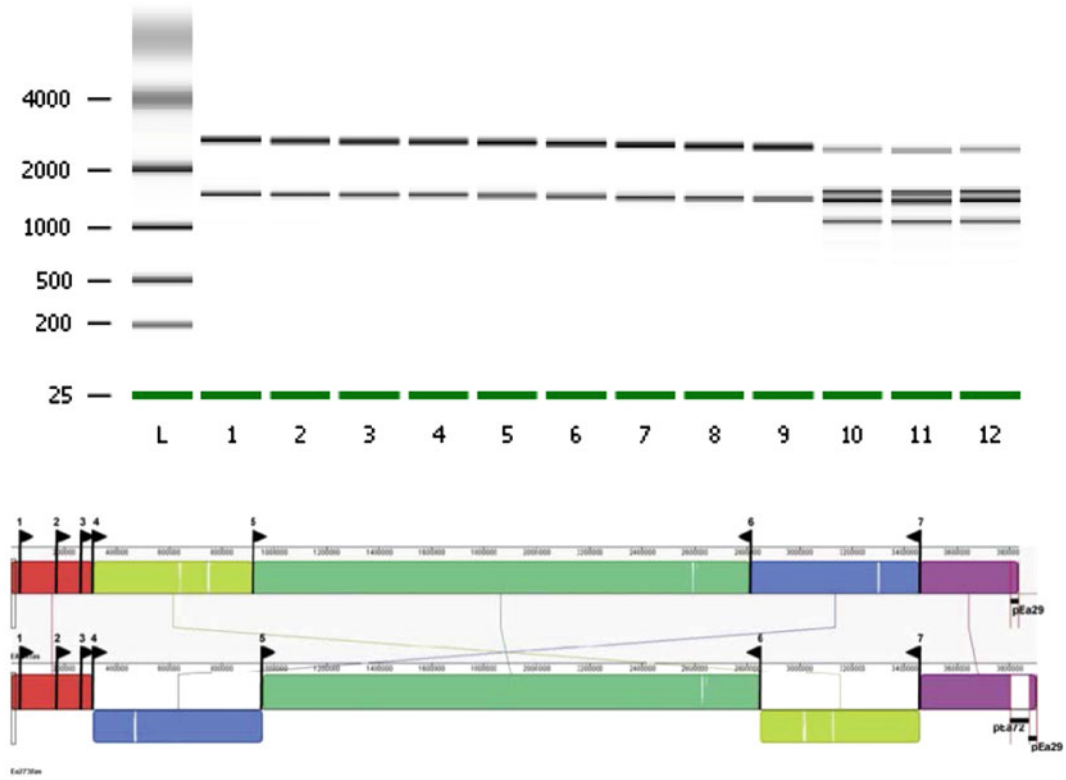

Fig. 1.3 *Top* Comparison of *E. coli* and *E. amylovora* RNAs showing the 5:2 ratio of 23S rRNA fragmentation pattern using the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). *Lanes 1 to 9 E. coli*, *Lanes 10 to 12 E. amylovora*. *23S rRNA* 2.9 kb, *16S rRNA* 1.5 kb, *two extra bands* 1.7 and 1.2 kb, respectively. *Below* Structure comparison of the genomes of *E. amylovora* CFBP 1430 and ATCC 49946, indicating genome re-arrangements (reprinted from Smits et al. 2010b). *Vertical lines* indicate ribosomal (r)RNA operons. Numbers indicate the rRNA operon numerator, as used in the alignments. Arrows indicate the transcriptional direction of the rRNA operons

probably resulting from homologous recombination events (Kube et al. 2010; Smits et al. 2010a, b, 2013). However, there is only one single large-scale re-arrangement of the genomes for *E. amylovora* strains CFBP 1430 and ATCC49946, which represents two independent recombination events, and may have occurred within copies of the ribosomal (r)RNA operon (Fig. 1.3). Analyses of the genome structures of the remaining seven *Spiraeoidea*-infecting strains reveal that genome structures for five strains (all from Europe) are identical to that of CFBP1430, and the other two strains (CFBP2585 from Ireland and Ea266 from Canada) are identical to that of ATCC49946 (Mann et al. 2013). In silico analyses revealed the exact fragment sizes, which classify *E. amylovora* CFBP 1430 as a Pt-3 genotype pattern and ATCC49946 as a Pt-4 genotype using pulsed-field gel electrophoresis (PFGE) analysis of *Xba*I- digested genomic DNA (Smits et al. 2010b; Zhang and Geider 1997). In addition, the ITS regions between the 16S and 23S rRNA genes of three operons of *E. amylovora* CFBP 1430 contain the transfer (t)RNA-Ala and tRNA-Ile genes, and four operons contain the tRNA-Glu genes. In contrast, the ITS regions between the 16S and 23S rRNA genes of four operons of the *E. amylovora* ATCC49946 genome are composed of tRNA-Ala/tRNA-Ile-genes and three operons of tRNA-Glu-genes, further suggesting that the recombination event

Table 1.3 Overview of plasmids in *Erwinia amylovora* and related *Erwinia* species associated with pome fruit trees

Plasmid	Species	Strain	Size (kb)	Host	Origin	Accession
pEA29	*E. amylovora*	All[a]	28.2	Various	Various	FN666576
pEA72	*E. amylovora*	ATCC49946	71.5	Apple	NY	FN666577
pEI70	*E. amylovora*	ACW56400	65.8	Pear	Switzerland	CP002951
pEL60	*E. amylovora*	Leb66	60	Apple	Lebanon	NC-005246
pEA34	*E. amylovora*	CA11	34	Apple	MI	NA
pEA30/ pEU30	*E. amylovora*	CFBP2585 UTRJ2	29.6	Sorbus Apple	Ireland Utah	HF560646 NC_005247
pEA8.7	*E. amylovora*	CA3R	8.7	Apple	CA	NA
pEAR5.2	*E. amylovora*	BAA-2158	5.25	*Rubus*	IL	FR719211
pEAR4.3	*E. amylovora*	BAA-2158	4.37	*Rubus*	IL	FR719210
pEA2.8	*E. amylovora*	IL-5	2.8	*Rubus*	IL	AY123047
pEA1.7	*E. amylovora*	IH3-1	1.7	*Crataegus*	LA	AY123046
pEP36	*E. pyrifoliae*	Ep1/96, 16/ 96	35.9	pear	Korea	FN392238
pEP5 pEP3 pEP2.6	*E. pyrifoliae*	Ep1/96, 16/ 96, Ejp617	4.96, 3.0, 2.6	pear	Korea Japan	FN392239 FN392237 FN392236
pJE01	*E. pyrifoliae*	Ejp617	30.9	pear	Japan	CP002125
pJE03	*E. pyrifoliae*	Ejp617	6.4	pear	Japan	CP002127
pET9, 35, 45, 46, 49	*E. tasmaniensis*	Et1/99	9.3, 35.4, 44.7, 46.1, 48.8	Apple flower	Australia	CU468128 CU468130 CU468131 CU468132 CU468133
pEb102, pEB170	*E. billingiae*	Eb661	102, 170	tree	UK	FP236826 FP236830
pEPIR37	*E. piriflorinigrans*	CFBP5887	37	pear	Spain	HE792893

Data adapted from Foster et al. (2004), Kube et al. (2008, 2010), Llop et al. (2011), Mann et al. (2012, 2013), Park et al. (2011), Powney et al. (2011), Sebaihia et al. (2010), Smits et al. (2010a, b, 2013)

[a] Except strain UPN527

may occur within the rRNA operon (Smits et al. 2010b). It is remarkable to discover how the *E. amylovora* genome structure has changed since the spread of the bacterium from North America to Europe about 60 years ago.

1.2.1.3 Plasmids

Comparative genomic analyses of several species of plant pathogenic bacteria revealed that the majority of strain-specific genes are plasmid-borne, indicating that acquisition and maintenance of plasmids may represent a major mechanism for bacteria to change their genetic composition and acquire new virulence factors (Zhao and Qi 2011). Plasmids also may contribute to genetic diversity of the *Spiraeoideae*-infecting strains of *E. amylovora* (Mann et al. 2013; Zhao and Qi 2011). A total of 11 plasmids have been reported in *E. amylovora* (Table 1.3) of which six are present among 12 sequenced *E. amylovora* genomes (Foster et al. 2004; Llop et al. 2011; Mann et al. 2013). The nearly ubiquitous plasmid pEA29 is present in all sequenced strains except UPN527 (Llop et al. 2006; Mann et al. 2013; McGhee and Jones 2000). Plasmid pEA29 contains genes encoding for thiamine biosynthesis, and loss of the *thiOSGF* genes results in thiamine auxotrophy (Llop et al. 2012; McGhee and Sundin 2008).

The major genetic distinction of *E. amylovora* strain ATCC49946 is the presence of plasmid pEA72, which is not present in the 11 *E. amylovora* genomes sequenced. Plasmid pEA72 contains a type IV secretion system that may be involved in conjugative transfer of the plasmid (Llop et al. 2012). In addition, plasmids pEA30, pEI70, and two small plasmids (pEAR5.2 and pEAR4.3) are present in strain CFBP 2585, ACW 56400, and ATCC BAA-2158, respectively (Table 1.3). In three sequenced *E. pyrifoliae* strains, three plasmids are common among them; and the fourth, pEp36, is not present in the Japanese strain, Ejp617 (Table 1.3). Instead, strain Ejp617 contains two extra plasmids (pJE01 and pJE03) (Kube et al. 2010; Park et al. 2011). Furthermore, *E. piriflorinigrans, E. billingiae,* and *E. tasmaniensis* strains contain one, two, and five plasmids, respectively (Table 1.3) (Kube et al. 2008, 2010; Smits et al. 2013).

1.2.2 Pan-Genome of *E. amylovora*

A pan-genome includes the full complement of genes in a species, which consists of the "core genome" containing genes present in all strains, a "dispensable or accessory genome" containing genes present in two or more strains, and finally "unique genes" specific to a single strain (Medini et al. 2005). It is predicted that the pan-genome of *E. amylovora* is still "open" based on two complete and 10 draft genome sequences (Smits et al. 2011; Mann et al. 2013).

1.2.2.1 Protein-Coding Genes, Core Genome, and Pan-Genome

The numbers of protein-coding genes in the genome of *E. amylovora* and related *Erwinia* species are listed in Tables 1.1 and 1.2. A comparison of genomes of *E. amylovora* strains CFBP1430 and ATCC 49946 shows that the two genomes share more than 99.9 % identity at the nucleotide level. However, based on initial annotation, the total predicted proteins in strain ATCC 49946 and CFBP1430 are 3565 and 3706, respectively (Sebaihia et al. 2010; Smits et al. 2010b). In a recent report, the number of predicted proteins for strain ATCC49946 is revised to 3712 (Mann et al. 2013). The numbers of predicted proteins in *Rubus*-infecting strains Ea644 and MR-1 are slightly higher as compared to those observed for *Spiraeoideae*-infecting strains of *E. amylovora* (Mann et al. 2013; Powney et al. 2011). Similarly, the genomes of the two *E. pyrifoliae* strains from Korea (Ep1/96 and DSM 12163 (Ep16/99) are almost identical; however, due to similar discrepancies in annotation, the total predicted proteins are 3697 and 4038 in Ep1/96 and DSM 12163, respectively (Kube et al. 2010; Smits et al. 2010a). On the other hand, *E. tasmaniensis* and *E. billingiae*, the two genetically most distant *Erwinia* species associated with pome fruit trees, contain 3,622 and 4,917 predicted proteins, respectively (Table 1.1) (Kube et al. 2008, 2010).

Using the subtractive hybridization-based mGenomeSubtractor program, which compares the reference genome against multiple bacterial genomes for in silico comparative genomic analyses, Zhao and Qi (2011) found that the number of conserved proteins with homology values greater than 0.81 is about 2,100. This indicates that the corresponding genes probably constitute the "core genome" among sequenced *E. amylovora* and related species (Zhao and Qi 2011). When *E. amylovora* strains CFBP1430 or ATCC 49946 are compared to ATCC BAA2158, a *Rubus*-infecting strain more closely related to the *Spiraeoideae*-infecting strains, more than 3,400 of the 3,500 conserved proteins (98 %) have homology values of 1. This indicates that the genomes of these *E. amylovora* strains are identical (Zhao and Qi 2011). When compared to all 12 sequenced *E. amylovora* genomes, including the genetically diverse *Rubus*-infecting strains (MR-1 and Ea644), about 3414 coding sequences are identified as core genes (Mann et al. 2013).

When *E. amylovora* strains are compared to sequenced *E. pyrifoliae, E. tasmaniensis,* and *E.*

billingiae strains, the numbers of conserved proteins are about 2,800, 2,600, and 2,200, respectively; and the number of proteins with homology values of 1 drops dramatically to 1,200 and below (Zhao and Qi 2011). This indicates that more diversification occurs for these pathogenic/saprophytic microorganisms and that *E. amylovora* and *E. pyrifoliae* may be evolutionally derived from two separate sources, one in North America and the other in Asia. Similar conclusions could also be drawn for strains of *E. pyrifoliae* from Japan and Korea, whereby about 85 % of conserved proteins (2,800 out of 3,300) are identical (Zhao and Qi 2011). In contrast, the number of strain-specific proteins varies among genomes (Zhao and Qi 2011). The majority of specific proteins among *Erwinia* species that have homology values of 0 are plasmid-borne, indicating that acquisition and maintenance of plasmids may represent a major mechanism for erwinias to change their genetic composition. The ever expanding pan-genome of *E. amylovora* is currently calculated to contain 5751 coding sequences based on 12 genome sequences (Mann et al. 2013; Smits et al. 2011).

1.2.2.2 Genomic Islands and Pathogenicity Islands

Genomic islands (GIs) are defined as clusters of genes in prokaryotic genomes, which may be acquired by horizontal gene transfer, and include prophages, integrated plasmids, integrative conjugative elements, integrons, and conjugative transposons (Langille et al. 2010). Typically, GIs contain mobility-related genes and may also carry "cargo" genes that can be involved in virulence, resistance, and ecological fitness (Seth-Smith and Croucher 2009). Pathogenicity islands (PAIs) are generally regarded as large regions of chromosomal or plasmid DNA containing multiple virulence genes, which are flanked by repeated sequences and are characteristically distinct in GC content from the rest of the genome (Hacker et al. 1997). Among the 12 sequenced *E. amylovora* strains, 12 GIs and three PAIs have been identified, and the former represent the majority of the genetic variation observed within the chromosomal component of the pan-genome (Mann et al. 2013; Zhao and Qi 2011). The majority of coding sequences within the GIs of the *E. amylovora* pan-genome consists of hypothetical and mobility-related genes, including genes involved in replication, transfer, and integration of mobile elements (Mann et al. 2013). In contrast, three T3SS genes and some effectors are located within the three PAIs of *E. amylovora* (Zhao et al. 2009a).

Among the GIs in the genome of *E. amylovora* strains, one 34.5-kb GI is present in the *Rubus*-infecting strains Ea644 and MR1, but a different GI of 23.4-kb occupies the same locus in the *Spiraeoideae*-infecting strains and ATCC BAA-2158 (Mann et al. 2013). Strains Ea644 and MR1 contain a type 1 restriction modification system, which protects the host DNA by adding methyl groups to recognition sites of expressed restriction enzymes. The *Spiraeoideae*-infecting strains encode a DNA degradation (Dnd) host-specific modification system, which incorporates sulfur into the DNA backbone to prevent restriction recognition (Mann et al. 2013). In addition, all three sequenced *Rubus*-infecting strains contain one 20-kb GI, which is absent in all nine sequenced *Spiraeoideae*-infecting strains. This locus encodes three polyketide synthases (PKS), a non-ribosomal peptide synthase (NRPS), and a putative transporter, which may represent a novel NRPS/PKS system for metabolite production (Mann et al. 2013). Remnants of this locus are found in CRISPR region 1 (CRR1) of the *Spiraeoideae*-infecting strains, suggesting that this GI in *Rubus*-infecting strains may be ancestral to CRR1 of the *Spiraeoideae*-infecting strains (see below) (Rezzonico et al. 2011).

There are large differences in the island transfer (IT) region between each of the *Rubus* strains and the *Spiraeoideae*-infecting strains (Mann et al. 2012, 2013). The IT region, an integrative conjugative element (ICE), is next to the *hrp* PAI1 HEE region (Oh and Beer 2005). The IT regions in *Spiraeoideae*-infecting strains are highly conserved, but the IT regions of the *Rubus*-infecting strains, *E. piriflorinigrans,* and

Table 1.4 Virulence-associated traits and their distribution in *E. amylovora* and related *Erwinia* species associated with pome fruit trees

Strains	*E. amylovora*			*E. pyrifoliae*			*E. tasmaniensis*	*E. billingiae*
Traits	CFBP 1430	ATCC 49946	BAA 2158	DSM 12163	EP 1/96	Ejp 617	Et1/99	Eb661
T3SS PAI1	+	+	+	+	+	+	+ (P)	–
T3SS PAI2	+	+	+	+	+	+	+	–
T3SS PAI3	+	+	+	–	–	–	+ (P)	–
Flagella 1 (S)	+	+	+	+	+	+	+	+
Flagella 2 (C)	+	+	+	+	+	+	–	–
Amylovoran biosynthesis[a]	+	+	+	+	+	+	+ (E)	+ (E)
Levansucrase (*lsc*)	+	+	+	–	–	–	+	+
Regulators of levansucrase (*rlsABC*)	+	+	+	+	+	+	+	–
Sorbitol metabolism (*srlAEBDMR*)	+	+	+	+	+	+	–	+
Protease A (*prtADEF*)	+	+	+	–	–	–	–	–
Siderophore biosynthesis (*dfoA*)	+	+	+	+	+	+	+	–
hopC1(hopPtoC)	+	+	+	–	–	–	–	–
hopAK1 (eop2)	+	+	+	–	–	–	–	–
hopX1 (eop3)[b]	+	+	+[b]	+	+	+	–	–
avrRpt2(eop4)[b]	+	+	+[b]	–	–	–	–	–
eop1/esc1[b]	+	+	+[b]	+	+	+	+	–
HrpK, HsvABC	+	+	+[b]	+	+	+	–	–

P partial, *S* separated, *C* clustered, *E* In Et1/99 and Eb661, the *amsE* gene is missing, but additional genes are present

[a] Some genes such as *amsCDE* are very diverse among different species of *Erwinia*

[b] Sequence diversification found in different species and even between *Spiraeoideae*- and *Rubus*-infecting strains. HrpK is truncated in strain ATCC BAA-2158. A single-base deletion at position 165 of AvrRpt2 found in strains Ea644 and MR-1; an amino acid substitution (C156S) of AvrRpt2 found in some *E. amylovora* strains isolated from Canada and US (Ea110, Ea400). Data adapted from: Kube et al. (2010), Mann et al. (2013), Vogt et al. (2013), Smits et al. (2010a, b, 2011), Zhao and Qi (2011)

E. pyrifoliae strains vary in gene content and length, showing a mosaic structure (Mann et al. 2012, 2013). It appears that the IT regions have undergone significant genome reduction in the *Spiraeoideae*-infecting strains, being more than 30 kb shorter in length than all sequenced *Rubus*-infecting strains (Mann et al. 2013). The IT region in *E. piriflorinigrans* is the largest (72 kb) (Smits et al. 2013).

Besides the well-characterized *hrp* T3SS PAI1, analysis of the complete genome sequence of *E. amylovora* strains revealed two extra non-flagellar T3SS PAIs (PAI2 and PAI3) and two flagellar T3SS systems (Flg-1 and Flg-2) (Table 1.4). PAI2 and PAI3 are 38 and 43 kb in length and contain 24 and 26 genes, respectively. Both PAI2 and PAI3 have a significantly lower %G+C content (38.4 and 43.4 % mol G+C, respectively) and similar gene organization with the known Inv/Spa-like T3SS PAIs of the insect endosymbiont *Sodalis glossinidius* and *ysa* of *Yersinia enterocolitica* (Zhao et al. 2009a). The non-flagellar T3SSs can be divided into at least five groups based on phylogenetic

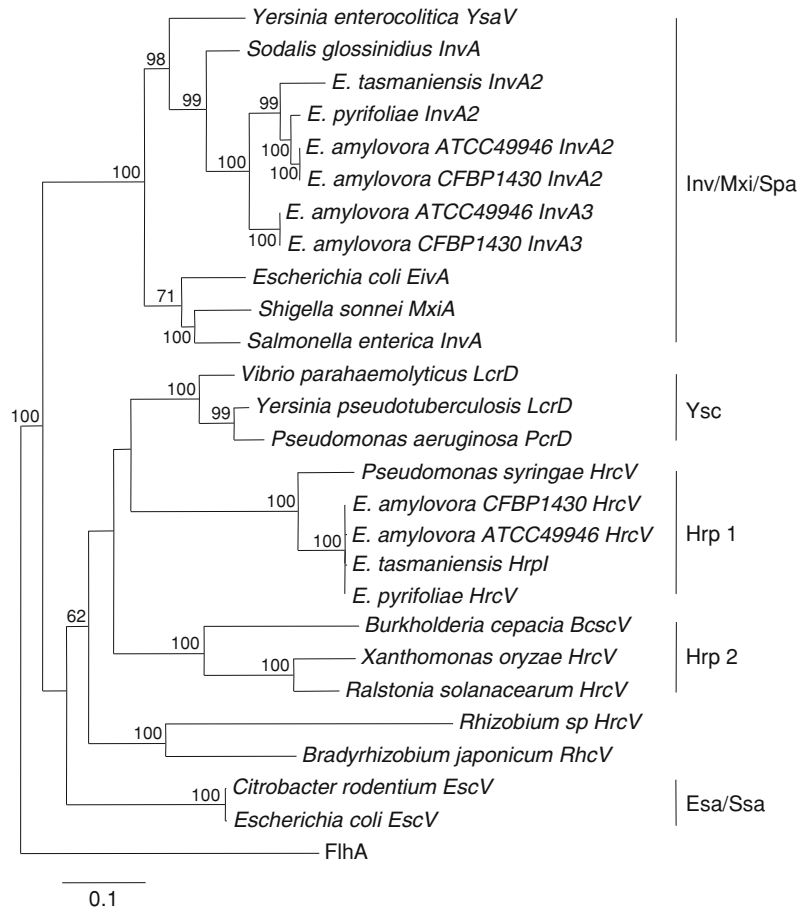

Fig. 1.4 Phylogenetic tree based on aligned amino acid sequences of HrcV/FlhA homologs from Zhao et al. (2011). The number at each node is bootstrap confidence values from 1,000 replicate maximum-likelihood trees. Similar copies of *invA* from *Erwinia* spp. (PAI2 and PAI3) are indicated by the number. The five T3SS groups are indicated on the *right*. FlhA was used as an outgroup

analysis (He et al. 2004). Phylogenetic trees based on the HrcV or InvA protein sequences revealed the PAI1 belongs to the Hrp1 group, whereas PAI2 and PAI3 belong to Inv/Mxi/Spa group (Fig. 1.4). The function of PAI2 and PAI3 is still unknown, but they are not directly involved in virulence to plants (Zhao et al. 2009a).

Analyses of genome sequences of closely related *Erwinia* species indicated that most PAIs are present in *E. piriflorinigrans, E. pyrifoliae,* and *E. tasmaniensis* (Table 1.4). However, PAI2 is not present in *E. billingiae* (Kube et al. 2008, 2010; Smits et al. 2010a, b, 2013), PAI3 is absent in the genome of *E. pyrifoliae* and *E. billingiae*, and only parts of PAI3 are present in *E. tasmaniensis* (Kube et al. 2008, 2010; Smits et al. 2010a, b). In addition, *E. pyrifoliae, E. piriflorinigrans,* and *E. tasmaniensis* strain Et1/99 contain *hrp* T3SS PAI1, but *E. piriflorinigrans* and *E. tasmaniensis* lack the HAE region that includes the effector *hrpK* and *hsvABC* genes (Kube et al. 2008; Smits et al. 2013). Furthermore, Flg-1 (separated into four gene clusters in the genome) is present in all *Erwinia* genomes; however, Flg-2 (clustered) is only present in *E. amylovora* and *E. pyrifoliae* strains (Smits et al. 2011; Zhao and Qi 2011).

1.2.2.3 Repetitive DNA and CRISPR Elements

Clustered regularly interspaced short palindromic repeats (CRISPR) represent a family of short DNA repeat sequences found in most archaeal and bacterial genomes (Horvath and Barrangou 2010). CRISPR typically consists of several non-contiguous direct repeats of 21–47 bp in size and

are separated by stretches of variable sequences (spacers). CRISPR is often adjacent to *cas* (CRISPR-associated) and *cse* (CRISPR Cascade complex) genes. Many of the spacer sequences associated with CRISPRs share sequence identity with bacteriophage, plasmid, and other laterally transferred DNA sequences. Thus, CRISPR/Cas systems function as CRISPR RNA (crRNA)-mediated adaptive immunity systems against bacteriophages and conjugative plasmids for sequence-specific detection and silencing of foreign DNAs, similar to RNA interference (RNAi) pathways in eukaryotes (Marraffini and Sontheimer 2010; Wiedenheft et al. 2012). In particular, Cas9, a DNA nuclease, has been shown to use dual-RNAs for site-specific DNA cleavage, which highlights the potential to exploit the system for RNA-programmable genome editing (Cho et al. 2013; Jinek et al. 2012).

Three *C*RISPR *r*epeat *r*egions (CRR1, CRR2, and CRR4) have been identified in genomes of *E. amylovora* and related species regardless of host range (McGhee and Sundin 2012; Rezzonico et al. 2011; Smits et al. 2010b). *E. pyrifoliae* strains contain four CRRs (CRR1 to CRR4), and CRR1, CRR2, and CRR4 repeats share 100 % sequence identity to those of CRRs in *E. amylovora*. *E. piriflorinigrans* and *E. tasmaniensis* contain two CRRs (CRR3 and CRR4). The repeats of both CRR1 and CRR2 are 29 bp in length, and only two nucleotide substitutions (GA to AT) at positions 14 and 15 differentiate CRR1 and CRR2 repeats (McGhee and Sundin 2012). The repeats of CRR3 and CRR4 are significantly different from those of CRR1 and CRR2 and are 28 bp in length. The CRR3 repeats of *E. pyrifoliae* and *E. tasmaniensis* are identical to each other, but differ by one bp from the CRR4 repeats of *E. amylovora* and *E. pyrifoliae* (Rezzonico et al. 2011).

In *E. amylovora*, the majority of spacers are characteristically 32 bp in length, ranging from 30 to 34 bp with some exception (Rezzonico et al. 2011). The number of spacers within CRR1 and CRR2 is variable among strains and ranges from 12 to 98 and 23 to 49 within CRR1 and CRR2, respectively (McGhee and Sundin 2012; Rezzonico et al. 2011). All strains contain five 34–35-bp spacers in CRR3 and CRR4. A total of 588 unique spacers have been identified in 85 *E. amylovora* strains (McGhee and Sundin 2012). Among them, approximately 23 % of the spacers match known sequences, including 16 % plasmids and 5 % bacteriophage. The plasmid pEU30, isolated from *E. amylovora* strains from the Western USA, is targeted by 55 spacers (Foster et al. 2004; McGhee and Sundin 2012). Interestingly, spacers from *E. pyrifoliae* do not share homology with those of *E. amylovora* (McGhee and Sundin 2012).

Both *E. amylovora* and *E. pyrifoliae* contain eight *E. coli*-type *cse* and *cas* genes between CRR1 and CRR2, and housekeeping genes between CRR2 or CRR3 and CRR4, respectively (Rezzonico et al. 2011). However, four housekeeping genes within a 2.56-kb region between the 3′ end of the CRR1 and the *cas3* gene is missing in *E. pyrifoliae, E. amylovora Rubus* strains, and some *E. amylovora* strains from the Western USA (CRISPR group III) (McGhee and Sundin 2012). In contrast, *E. pyrifoliae, E. piriflorinigrans,* and *E. tasmaniensis* have *Y. pestis* subtype *csy* genes between CRR3 and CRR4. The *csy* genes and CRR3 are apparently lost in *E. amylovora*, leaving only CRR4 as a relic (Rezzonico et al. 2011).

1.2.3 Genomic Resources

The EMBL/GenBank accession numbers for *E. amylovora* genome and plasmid sequences are listed in Tables 1.1, 1.2 and 1.3. The *E. amylovora* microarray design and HrpL regulon microarray data are available at ArrayExpress Web site (http://www.ebi.ac.uk/arrayexpress/; accessions: Microarray #A-MEXP-2000, and Dataset #E-TABM-1137) (McNally et al. 2012). Other microarray data are available at the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/) (accession numbers GSE30751 (RcsBC regulons), GSE28630, GSE37064 (YbjN/AmyR regulons), and GSE45602 (T3SS inhibitors) (Wang et al. 2011a, 2012a, b; Yang et al. 2014).

1.3 Genome-Enabled Understanding of *E. amylovora* Virulence

During the past two decades, research on fire blight has made great strides in elucidating the genetic, molecular, and physiological basis for pathogenesis (Oh and Beer 2005). However, many key questions remain to be answered due to insufficient genetic information regarding *E. amylovora* and related *Erwinia* species, despite extensive studies on *E. amylovora* and fire blight disease for more than a century. Two puzzling questions for the fire blight research community are why natural isolates of *E. amylovora* display differential virulence and what are the molecular mechanisms underlying the host specificity of *Erwinia* strains (Zhao and Qi 2011). The recent revelation of the genetic composition of these microorganisms provides opportunities to employ both comparative and functional genomic approaches to understand the pathogen, its ability to cause disease, and its interaction between host plants and insect vectors.

1.3.1 Evolutionary Insights into Genome Sequencing of *E. amylovora* and Related Species

In prokaryotes, small subunit ribosomal RNA (rRNA) genes, especially 16S rRNA, have been universal phylogenetic markers for reconstructing the evolutionary relationships between microorganisms. Relatedness among closely related organisms can also be distinguished by analyzing the phylogenetic relationship of housekeeping genes (Fig. 1.1) (Sarkar and Guttman 2004; Zhao and Qi 2011). A global phylogeny of 191 fully sequenced organisms, representing the three domains of life, has been reconstructed based on a concatenated alignment of 31 single universal protein families related to ribosomal function (Ciccarelli et al. 2006). Phylogenomic trees have also been generated from more specialized data sets. A phylogenetic tree based on 14 flagellar proteins is congruent with a bacterial species tree using 25 single-copy proteins present in 249 genomes, indicating core components of the bacterial flagella may originate from a single ancestral gene through duplication and diversification (Liu and Ochman 2007). Furthermore, phylogenetic trees reconstructed from a concatenation of the seven core set of two-component systems from enterobacteria agreed well with that of the 16S rRNA gene (Qi et al. 2010).

1.3.1.1 Comparative Genomics and Evolution of *E. amylovora* and Related Species

In order to identify genes that are responsible for basic biology, virulence, and evolution of *E. amylovora* and its related species, and to determine the molecular mechanisms of the differential virulence and host specificity, comparisons of the complete genomes of *E. amylovora* strains and related *Erwinia* species have been performed (Kube et al. 2010; Mann et al. 2012, 2013; De Maayer et al. 2011; Rezzonico et al. 2012; Smits et al. 2010a, b, 2011, 2013; Thapa et al. 2013; Zhao and Qi 2011). These comparative genomic studies provided a preliminary scientific basis for determining the relatedness and evolution of genes/proteins within the genomes of *E. amylovora* and closely related *Erwinia* species (Smits et al. 2011; Zhao and Qi 2011).

Based on comparative genomic studies, a hypothesis of an evolutionary history of genome-sequenced *E. amylovora* strains and related *Erwinia* species has been proposed (Mann et al. 2012; Rezzonico et al. 2012; Smits et al. 2011). According to this hypothesis, the genealogy within the genus *Erwinia* agrees largely with the phylogeny (Fig. 1.1). From the ancestral *Erwinia* in the enterobacterial ancestor, it takes several evolutionary steps to reach the common ancestor for the genome-sequenced *Erwinia* species, which separate from other *Erwinia* spp. From there, the non-pathogenic *E. tasmaniensis* separates from the pathogenic *Erwinia* ancestor, which further differentiates into *E. piriflorinigrans, E. amylovora,* and *E. pyrifoliae* (Smits et al. 2010a). This hypothesis is supported by

specific features or traits discovered in the comparative genomic analyses (Malhony et al. 2012; Smits et al. 2010a, b). Most of the specific features or traits are summarized in Table 1.4 and some discussed below in detail.

Comparative genomic studies have identified the following ancestral origins of virulence-associated traits, including three T3SS PAIs (PAI1 to PAI3) (Mann et al. 2012; Zhao et al. 2009a), three type VI secretion systems (T6SS) (De Maayer et al. 2011), two flagellar systems (Zhao et al. 2011), one type I secretion system (Palacio-Bielsa et al. 2012), CRISPR repeat sequences and associated *cas*/*cse* genes (McGhee and Sundin 2012; Rezzonico et al. 2011), lipopolysaccharide, amylovoran and levan biosynthesis genes (Rezzonico et al. 2012; Smits et al. 2011), and genes involved in sorbitol metabolism (Mann et al. 2013; Smits et al. 2011). The major difference within the *hrp*-T3SS PAI1 between pathogenic *Erwinia* species and *E. tasmaniensis* is the IT and HAE regions; however, gene sequence diversification is also found within the HEE region as exemplified by the *eop1* gene among *E. amylovora* strains with different host ranges (Table 1.4).

Type VI secretion system (T6SS) has been identified in many Gram-negative bacteria (Records 2011); however, its role in virulence has not been documented in *E. amylovora*. In *E. amylovora*, three T6SS gene clusters (1–3) have been identified (De Maayer et al. 2011; Smits et al. 2010b). Variation between strains of *E. amylovora* is primarily found within the non-conserved *hcp* and *vgrG* islands of T6SS-1 regions II and IV, and T6SS-3 region IV (Mann et al. 2013). T6SS-1 and T6SS-2 are highly similar to the T6SS clusters of *E. pyrifoliae*, *E. piriflorinigrans, E. tasmaniensis*, and *E. billingiae* and some differences in the genes encoding VgrG proteins in T6SS-1. T6SS-2 also shows variations in gene content (Kube et al. 2010; Smits et al. 2010a), where a frame shift is found in one of the genes in *E. pyrifoliae* and a gene is lost in the corresponding cluster in *E. tasmaniensis* (Smits et al. 2010a, b). T6SS-3 is only identified in *E. amylovora*, but absent in other related *Erwinia* species (Smits et al. 2010b, 2013).

Lipopolysaccharides (LPS) are major components of the cell surface of Gram-negative bacteria. An LPS biosynthesis gene (*waaL*) in *Spiraeoideae*-infecting strains of *E. amylovora* has been shown to be involved in virulence (Berry et al. 2009). Based on genome sequences, *Spiraeoideae*-infecting strains contain three glycosyltransferases and an LPS ligase (*Spiraeoideae*-type *waaL*), whereas only two glycosyltransferases and a different LPS ligase (*Rubus*-type *waaL*) for *Rubus*-infecting strains exist in the core region of the LPS biosynthetic gene cluster (Rezzonico et al. 2012). These coding sequences for LPS biosynthesis genes share little to no homology at the amino acid level between *Rubus*- and *Spiraeoideae*-infecting strains (Mann et al. 2013). In addition, *amsCDE* within the *ams* biosynthetic operon are very diverse among different species of *Erwinia* (Langlotz et al. 2011).

Other genes seem to have been acquired after divergence of pathogenic species, including a second flagellar T3SS (Flg-2). Phylogenetic trees based on concatenation of 14 conserved flagellar proteins showed that both Flg-1 and Flg-2 are clustered with those of enterobacteria, indicating that these flagellar systems may be originated from other enterobacteria (Fig. 1.5) (Zhao et al. 2011). The Flg-1 system is much closer to the phylogeny of species trees than that of Flg-2, which is closely related to those of *Sodalis glossinidius*, suggesting that Flg-2 along with PAI2 and PAI3 may be acquired from a similar enterobacterial source by horizontal gene transfer (Figs. 1.4 and 1.5). Interestingly, PAI2 and PAI3 are also clustered together and closely related to those of *S. glossinidius* (Fig. 1.4) (Zhao et al. 2011).

Virulence factors such as type III effectors (Eop2, HopC1 and AvrRpt2) are present in *E. amylovora* strains, but not in *E. pyrifoliae* strains, and these effectors may contribute to host specificity (Khan et al. 2012; Zhao et al. 2005, 2006). The EPS levan is a virulence factor in *E. amylovora*; however, the levansucrase gene (*lsc*) is absent in the genome of *E. pyrifoliae* strains, which can be used to differentiate *E. amylovora* from *E. pyrifoliae* (Zhao and Qi