Phenotypic and Genetic diversity of *Agrobacterium tumefaciens* from walnut orchards in California

By

Limin Chen

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Plant Pathology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Richard M. Bostock, Chair

Daniel A. Kluepfel

Robert L. Gilbertson

Committee in Charge

2021

Acknowledgements

Finishing the four-year Ph.D. study at UC Davis was an amazing journey in my life. I have received a great deal of support and assistance from the university community. I would first like to express my deepest appreciation to my supervisor, Daniel Kluepfel. He provided me with this chance to pursue my Doctorate degree. He was very supportive of my interest in the world of genetic diversity of *A. tumefaciens*, and he encouraged me to keep learning.

I would like to acknowledge Amisha Poret-Peterson. She taught me bioinformatics and always offered me essential insight and feedback. I would also like to thank for my lab manager, Ali McClean. His top-quality wet lab experience has proven instrumental in my growth as a lab scientist. He is also a good friend and mentor. He always helped me to see how much progress I have made and encouraged me to never give up. The Data Analysis workshop gave me great supports on my bioinformatics work.

My QE committee guided me in editing the scope and direction of the dissertation. Without going through this process, I wouldn't have been able to understand how to prepare a reasonable proposal as a graduate student. Also, I would like to thank the dissertation committee. This committee was so important in the curation of my dissertation. I would like to acknowledge all professors in the Plant Pathology department and other programs. PLP is an amazing program that taught me not just knowledge on plant pathology, but also how to be a scientist and how to think critically. Everyone is so supportive to graduate students, which has been a positive influence on me.

In the end, thanks and appreciation are also extended to my family. Their love, caring and sacrifice taught me to be a good person, to be independent, and to have strong responsibility. Also, I express my thanks to my friends, Yan Lu and Joe Turkovich. Our more than 10 years of friendship has brought me light whenever I feel down; Joe's neighborly caring and support made Davis feel a bit like home. His welcome and support makes me feel the warmth and comfort of home. His caring is heart-touching which means a lot to me as an international student here.

Phenotypic and Genetic diversity of *Agrobacterium tumefaciens* from walnut orchards in California

Abstract

Crown gall (CG) caused by a Gram-negative bacterium *Agrobacterium tumefaciens* is a serious threat to the California walnut industry. Understanding the classification and diversity of this pathogen is important for the development of sustainable management strategies for CG. In the present study, average nucleotide identity (ANI), phylogenetic, and pangenomic analyses of 311 *Agrobacterium* strains, including 29 newly isolated and sequenced strains from walnut orchards in California were performed. These analyses revealed 35 ANI groups (genomic species) and four main clades corresponding to three previously described biovars, i.e., Biovars 1, 2, and 3 and a mix of known and unknown *Agrobacterium* species. The *Agrobacterium* genus had an open pangenome, indicating great diversity among individuals in this group. Five Phenotypic traits (virulence, growth, antibiotic resistance, K84 sensitivity, motility) of the 29 strains were determined. Virulence testing on *Datura stramonium* and walnut hybrid genotypes (*J. regia X J. hindsii*) revealed considerable variation. The growth rates ranged from 0.15 min^{-1} to 0.65 min^{-1} in a nutrient rich medium, whereas the growth rates in the nutrient poor medium ranged from 0.05 min^{-1} to 0.07 min^{-1}. Most strains were resistant to streptomycin and vancomycin, whereas they were sensitive to the other six tested antibiotics. Nearly half of the strains were resistant to the biocontrol *Agrobacterium* strain K84. Comparative genomic analysis of the 29 strains revealed a highly conserved T4SS, the T-DNA, and the *vir*A/*vir*G two component regulatory system. The 29 strains were divided into agropine-, succinamopine-, and nopaline-types. The agropine and succinamopine opine types were related to the size of tumors induced on *Datura stramonium*, but not on walnut. The different hosts (*D. stramonium* and walnut) also exerted a significant influence on the tumor size induced by a given *A. tumefaciens* opine genotype. These data indicate that the opine type may influence a strain's host preference. Interestingly, three virulent strains did not contain a T6SS. The remaining strains contained a classical T6SS in which the *imp* operon was conserved, whereas the *hcp*

operons were found to be variable among the strains. Variability in the *hcp* operon may contribute to the ability of *A. tumefaciens* to adapt to a wide range of environments. Collectively, these data will be useful in the development of effective crown gall management strategies for the walnut industry in California.

Table of Contents

Chapter I

Introduction

Crown Gall Disease on Walnut

In California, walnuts were first cultivated in the late 1700s by Franciscan Fathers [1]. In 1867, the orchardist and nurseryman, Joseph Sexton, started growing English walnuts in Santa Barbara County, California. This marked the beginning of commercial walnut plantings. Today, the California walnut industry is the largest exporter of walnuts in the world [2]. California growers produced approximately 676 thousand tons of walnuts in shell in 2018 [3], generating $1.4 billion in farm gate revenue and supporting some 60,000 jobs directly and indirectly. However, the walnut industry is impacted by several important soil-borne diseases, which inflict an estimated annual yield loss of $241 million. Consequently, the California Walnut Board has cited soil-borne diseases as major threats to the industry [4].

Crown gall (CG) is one of the major soil-borne diseases affecting walnut production in California [5]. The causal agent of CG is the Gram-negative bacterium *Agrobacterium tumefaciens,* which possesses one of the widest host ranges of any known plant pathogen, i.e., nearly all dicots are susceptible [6]. The main symptom of CG disease of walnut is the formation of tumors at the crown of the walnut rootstock. Currently, about 85% of the walnut industry uses the hybrid rootstock *Juglans hindsii* X *J. regia* [4], which is highly susceptible to *A. tumefaciens*.

Management of CG has long been a challenge all over the world due to its wide host range and unique pathogenesis mechanism [7]. The biocontrol strain *A. radiobacter* K84 [8] showed promise in managing of crown gall when it was first reported in 1972 by New & Kerr [9] and has been used to manage crown gall in many countries [10,11,12]. Strain K84 produces the bacteriocin named agrocin84, which inhibits DNA or RNA synthesis of sensitive virulent *A. tumefaciens* strains. Agrocin84 is encoded by genes located on the plasmid pAgK84 [13]. However, pAgK84 can be acquired by virulent strains, causing these strains to become resistant to the biocontrol strain K84 [14]. To solve this problem, a Tra- deletion mutant of pAgK84 was constructed and shown to significantly reduce pAGK84 transfer to virulent strains while still reducing gall formation on susceptible hosts [15].

Horizontal gene transfer of pAgK84 is one mechanism whereby virulent *A. tumefaciens* strains become resistant to biocontrol strain K84. Characterization of opine catabolism genes located in the *acc* operon indicate that mutations in the *accF* gene also lead to resistance to strain K84 [16]. The *acc* operon consists of 8 genes, named *accR*, and *accA* through *accG*, whose expression is induced by agrocinopines [17]. Theoretically, strains harboring the agrocinopine synthase gene are sensitive to K84 due to mimicry of agrocinopine by agrocin84 which enables its uptake by virulent strains. Thus, the virulent strains will become sensitive to strain K84 [18]. However, studies on strain K84 efficacy across the walnut growing region of California showed that K84 resistant strains are widely present. However, the resistance mechanism of these strains is unclear [19,20]. To date, there is no widely effective bactericide available on the market for post-plant control of crown gall of walnut. Thus, crown gall disease management remains a serious problem in walnut orchard in CA.

Phylogeny and classification of *A. tumefaciens* strains

The classification of *Agrobacterium* spp. has long been debated [21,22,23]. Conventionally, the *Agrobacterium* genus is grouped into three biovars, *A. tumefaciens* (biovar 1), *A. rhizogenes* (biovar 2), and *A. vitis* (biovar 3) [24,25]. This classification is based on biochemical tests, serology, pathogenicity, and protein patterns following electrophoresis [24]. However, the horizontally transferrable Ti plasmids make classification a challenge [26]. Thus, the biovar concept of *Agrobacterium* is first adopted at the intrasubspecific level, i.e., three biotypes (biovars) were detected of 53 pathogenic strains of *Agrobacterium radiobacter* var. *tumefaciens* [27], whereas these biovars are currently used as species level designations for pathogenic populations [28]. The genus *Agrobacterium* contains many plant pathogenic species, such as *A. larrymoorei*, *A. rhizogenes*, *A.rubi*, *A. tumefaciens,* and *A.vitis* [29]. Young et al. [23] in 2001 proposed placement of *Agrobacterium* into the genus *Rhizobium*. The authors argued that the 16S rRNA sequence data was unable to resolve *Agrobacterium* spp. from *Rhizobium* spp.[30,31]. However, Farrand et al. in 2003 [21] proposed that the genus *Agrobacterium* should be retained based on restriction fragments

length polymorphism (RFLP) analyses. As sequencing technology became more affordable, single copy housekeeping genes have been employed for *Agrobacterium* spp. classification. For example, genotyping results of *Agrobacterium* spp. using the *recA* gene sequence data are in total agreement with the results obtained using amplified fragment length polymorphism (AFLP)-based analysis[32]. As sequence data for more housekeeping genes became available, phylogenies built from various housekeeping genes (e.g. *recA, mutS, gyrB, glgC* and *gltD*) have been used for *Agrobacterium* spp. classification [32,33].

A Bayesian phylogeny constructed from four concatenated housekeeping genes (*rrs-atpD-recA-rpoB*) indicates that *Agrobacterium* genus exists as a monophyletic group [34]. Weisberg et al., [35] studied the evolutionary relationship of Rhizobiaceae using a multilocus sequence (MLS) phylogeny constructed from 24 conserved genes, and identified *Agrobacteria* lineages that evolved independently from rhizobia lineages. Taxonomy inferred with MLS phylogenetic analyses have gradually been embraced by the *Agrobacterium* community [22]. A recent review on *Agrobacterium* taxonomy concluded that the genus *Agrobacterium* contains 14 species, and suggests *A. vitis* and *A. rhizogenes* should be moved into the genus *Allorhizobium* and *Rhizobium,* respectively [36]. However, when comparing various MLS-based phylogenies, the taxonomy of *Agrobacterium* and Rhizobiaceae is still controversial. More accurate and comprehensive information is required to improve the resolution of *Agrobacterium* taxonomy. In my research, I follow the classification approach, in which the *Agrobacterium* genus includes the species *A. tumefaciens*, *A. vitis*, and *A. rhizogenes*.

The current consensus concept of a bacterial species is defined as a group of strains, which show greater than 70% DNA-DNA hybridization (DDH) [37]. This 70% DDH standard has been shown to correspond with an ~95% average nucleotide identity (ANI) [38,39]. Phylogenomic analysis has become a popular tool for the classification of bacteria at both the genus and species levels [40,41]. Phenotypically similar, but genotypically distinct groups of strains have been referred to as genomic species, genomic groups, genomospecies, or genomovars [42,43]. Homologous recombination studies in *Agrobacterium* have indicated

the genomic species concept agrees with the species concept in bacteria described above [44]. *A. tumefaciens* was further divided into 10 genomic species G1-G9 and G13 based on homologous recombination results [45]. However, it is not known if the genomic species concept will concur with the species concept based on ANI value (> 95% identity). In my thesis research, I explore the use of ANI values to characterize *Agrobacterium* genomes.

In parallel with ANI values being used to facilitate species- or genus-level classification, the Core Genome Hypothesis (CGH) can be used to define variation within a population [46]. CGH is proposed to define species-specific phenotypic clusters [47]. In 2005, the pangenome concept, first created by Tettelin et al. (2005), defined the sum of the core genome (all conserved genes), the dispensable genome (accessory genes) and unique genes specific to a given strain [48,49]. Pangenome analysis has become increasingly popular as a method to reveal important genetic elements involved in pathogenesis, antibiotic resistance, etc. [48,50]. Pangenome analysis also has been applied to the study of the evolution of host-microbe interactions and niche adaptation for Rhizobiales [51]. The analysis successfully distinguished strains of *Agrobacterium* and *Rhizobium* from those of other genera in the family [51]. The presence/absence gene profile obtained in this *Rhizobium* pangenome analysis confirmed the phylogenomic pattern of species divergence [52]. However, applying pangenome analysis to uncover *Agrobacterium* spp. taxonomy has not been reported.

Both whole genome ANI value calculations and pangenome analyses require high-quality genomic sequence data. Fortunately, next-generation sequencing (NGS) and third-generation sequencing technologies have reduced the sequencing costs while generating high quality genomes [53]. The 29 complete assembled *A. tumefaciens* genomes presented in my thesis, along with access to the publicly available draft and complete genomes of 282 *A. tumefaciens* strains facilitated our comprehensive evaluation of the genus *Agrobacterium* to provide a firm basis for its classification [22].

Virulence mechanism of crown gall disease

The tumor-inducing plasmid (pTi) is required for virulence of *A. tumefaciens* [54]. The loss of pTi results in loss of pathogenicity [54]. The Ti plasmid harbors genes involved in replication, conjugation, and opine catabolism [54]. In addition, the Ti plasmid contains a virulence (*vir*) region that contains genes encoding for type IV secretion system, other virulence effector proteins, and the transfer DNA or T-DNA region. The T-DNA possesses two group of genes: oncogenes and opine biosynthetic genes, which are expressed once the T-DNA is inserted into the host genome. Expression of these genes leads to the development of galls or tumors, i.e. the crown gall phenotype [55]. Ti plasmids are divided into groups based on the types of opines they produce, such as octopine, nopaline, agropine, mannopine, etc. These opines are the products of conjugation of amino acids with ketoacids or sugars, and are thought to serve as a source of carbon and nitrogen for *A. tumefaciens* [18].

The pathogenesis process of *A. tumefaciens* has been studied extensively [22]. Motility, chemotaxis, and attachment are all important pre-infection processes of virulent *A. tumefaciens* strains [56]. The rhizosphere is rich in sugars, amino acids, and other compounds, which serve as signal molecules for *A. tumefaciens* in this pre-infection process. In addition, plant wound exudates contain the signal molecule, acetosyringone (AS), which can be sensed by the VirA/VirG two components regulatory system of *A. tumefaciens* strains. VirA is a transmembrane histidine kinase, and has a C-terminal cytoplasmic domain that is autophosphorylated by AS. The phosphorylated VirA then activates VirG protein in the cytoplasm. The VirG protein, which is a response regulator, modulates downstream signal transmission and gene expression. The VirA/VirG two component regulatory system performs a multifunctional role. When acetosyringone concentration is low, it mediates chemotaxis, whereas when concentrations are high, it initiates expression of the *vir* operon [57].

The *vir* operon contains genes encoding a type IV secretion system responsible for T-DNA transfer to plant cells where it is stably inserted into the plant genome [22, 58]. Once the T-DNA is inserted into the

plant genome, oncogenes and opine biosynthetic genes in the T-DNA are expressed by the plant transcription and translation system [22, 58]. Plant hormones encoded by oncogenes cause uncontrolled proliferation of plant cells, which results in tumor formation. At the same time, products of opine biosynthesis genes produce opines, which serve as carbon and nitrogen sources for *A. tumefaciens* strains harboring opine-specific catabolism genes on the same Ti plasmid [22, 58, 59]. Opines also serve as inducers to promote conjugation of Ti plasmids among agrobacteria, resulting in additional *A. tumefaciens* strains being able to take advantage of available opines [60].

In comparing different opine type genes, Ti and root-inducing (Ri) plasmids shared homologous DNA regions, which mediate virulence [61]. The homologous regions contain essential genes for pathogenesis, e.g., *virD1, virD2, virB* genes of the T4SS*, virE2*; *virC*; and *virC2*, all of which are necessary for transformation. In contrast, nonessential factors such as *virD3, virD5, virE3, virF, virH, virJ, virK, virL,* and *virM* are variable among different strains [62]. These essential and nonessential genes have been further defined based on examination of *A. tumefaciens* mutants introduced into wounded susceptible hosts [62]. In addition, chromosomes and pAt are known to carry nonessential genes involved in *A. tumefaciens* pathogenesis, such as *chvE, chvH, and chvI* [62]. Some of these genes are also speculated to mediate host adaptation [62]. It is also reasonable to infer that these nonessential genetic elements may contribute the observed variability in virulence observed for *A. tumefaciens* strains in nature.

Crown gall formation involves the interaction between genes from both virulent *A. tumefaciens* strains and susceptible host plants [62]. Although pTi carries key genes that mediate crown gall, other virulence genes are found on the chromosomes, such as *chv* genes [63]. *A. tumefaciens* strains with mutations in *ChvA* and *ChvB* are nonpathogenic because they are unable to attach to plant cells [64,65]. A *ChvD* mutant has been described that is attenuated in *virG* gene expression and attenuated in virulence [66]. Mutation of *ChvH* led to reduced virulence because of decreased expression of many vir genes, including *VirB9, VirB10, VirB11*, *VirG* and *VirE*. [67]. ChvG/I is a two-component regulatory system induced

by low pH, which activates expression of the transcriptional regulator VirG [68]. The ChvE protein has a role in VirA/G signaling and binds to sugar transporter as well as chemotaxis proteins [69,70,71]. However, only a few model strains (e.g., C58 and A6) have been used to characterize these pathogenesis mechanisms [62]. This provides a limited view into the genetic diversity of virulent *A. tumefaciens* strains. It is important to gain comprehensive knowledge of the underlying genetic basis of virulence by performing comparative genomic analysis on a range of genetically diverse *A. tumefaciens* strains.

Type VI secretion system (T6SS)

Gram-negative bacteria have evolved various remarkable and sophisticated secretion systems to secret either proteins or DNA molecules into the environment. The well-studied secretion systems include types I-VI[72]. Many phytopathogenic Gram-negative bacteria possess a type III secretion system (T3SS), which secretes effectors into the host cell cytoplasm. Many species of the genera *Pseudomonas*, *Ralstonia*, *Erwinia*, and *Xanthomonas* contain well characterized T3SSs [72]. The T3SS serves as an essential virulence mechanism for many bacterial pathogens infecting plant hosts [73]. Virulent *A. tumefaciens* strains possess a T4SS, which exports the T-DNA molecules and protein effectors into the plant host and lead to tumor formation [74]. The more recently characterized T6SS is also present in *A. tumefaciens* strains [75].

However, little is known concerning the distribution and genetic diversity of T6SS in *A. tumefaciens* strains, mainly those affecting the walnut industry of California. The T6SS was first characterized as a secretion apparatus in *Pseudomonas aeruginosa,* where it was known to export the hemolysin co-regulation protein (Hcp1). This secretion system was later found in more than 25% of Gram-negative bacterial genomes [76]. This widely conserved multicomponent nanomachine is structurally related to contractile phage tails [77,78]. It is one of the many specialized secretion systems that bacteria use to transport proteins and other factors into the environment and other microbes [72]. The T6SS can function either in a contact-dependent [79,80] or contact-independent manner to compete with other microbes [81,82]. In the model *A. tumefaciens* strain C58, the T6SS-encoding region is comprised of two divergently

transcribed operons. One is the *imp* operon consisting of 14 genes (*atu4343* to *atu4330*) [83], eleven of which are considered to be core for the T6SS in proteobacteria [75]. The other is the *hcp* operon encoding 9 genes (*atu4344* to *atu4352*), four of which are core for the T6SS in proteobacteria[84].

One study evaluating the T6SS of eleven *A. tumefaciens* strains showed that the imp loci are conserved, whereas genes in the *hcp* operon are more variable, including *vgrG* [85]. The 11 strains represent 4 genomic species of *A. tumefaciens* and 6 different types of *hcp* operons. In addition, *vgrG2* locus 2, *vgrG* locus 3 and *vgrG* locus 4 are also present in some *Agrobacterium* strains [85]. Because T6SS is important for bacteria to compete for ecological niches, the diversity of T6SSs in eleven strains isolated from various sources suggested that the T6SS may facilitate host adaptation of strains to a given environment. This study provided evidence of substantial genetic diversity of T6SS among *A. tumefaciens* strains.

Antibiotics and antibiotic resistance

The use of antibiotics in agriculture is largely restricted to high-value fruit and vegetable crops and is registered in few countries. For example, streptomycin is registered in the USA, Mexico, Canada, New Zealand, Germany, Switzerland, and Austria. The real data for application of antibiotics on plants and crops worldwide is unknown [86]. In the USA, there has been more than 50 years of antibiotic use for crop protection [86]. Antibiotics usually are used to prevent plant pathogenic bacteria from growing. Once the symptoms are visible, it is often too late to treat with antibiotics. Also, the efficacy of antibiotic applications last less than a week. This results in the high frequency use of antibiotics in plant protection [86].

Streptomycin is one of the most commonly used antibiotics for managing several bacterial diseases of plants [87,88]. Streptomycin is primarily used for controlling fire blight of pome fruits, e.g., pear and apple trees, and related ornamental plants [89]. Fire blight, caused by *Erwinia amylovora*, was first reported in New York and has spread to other parts of North America, New Zealand, the Middle East, and Europe [89]. This disease causes an annual loss of over $100 million for growers in the USA alone [89].

Streptomycin can be effective in managing this disease. However, *Erwinia* spp. resistant to streptomycin were detected 5 to 10 years after commercial use of this antibiotic [90,91]. Oxytetracycline is also used to manage bacterial spot of almonds caused by *Xanthomonas arboricola* pv. *pruni* and to manage walnut blight caused by *Xanthomonas arboricola* pv. *juglandis* [92].

Development of antibiotic-resistant bacteria is a great concern today in agriculture [93]. Antibiotic resistance genes can be acquired by plant bacterial pathogens from either the environment or from non-pathogens through horizontal gene transfer [94]. It is important to investigate how the antibiotic resistance genes are spread in agriculture systems, especially in economically important plant pathogens [95]. Thus, understanding the level and diversity of antibiotic resistance in *A. tumefaciens* population will facilitate the management of antibiotic usage. In my thesis, the antibiotic resistance profile of *A. tumefaciens* strains from walnut orchards in CA were established.

Growth Rate and Motility

Growth rate and motility are two important phenotypic traits affecting bacterial survival. The growth rate will determine pathogen population size in a specific period of time, which will influence inoculum load, competition, and ultimately disease incidence and severity. Bacterial motility facilitates host surface colonization, which is an important early step of pathogenesis [96]. Studies have shown that both laboratory and field investigations are necessary to gain understanding of bacterial movement in soils. Generally, quantitative comparisons from laboratory conditions reflect the field situations [97]. It is important to examine growth rates and motility on selected *A. tumefaciens* from walnut orchards, as this will improve our understanding of the phenotypic diversity of these *A. tumefaciens* strains.

Summary

Agrobacterium tumefaciens diversity across the walnut growing regions of the Central Valley of CA has not been characterized. Understanding *A. tumefaciens* biology and diversity of pathogenesis across this geographic scale will provide information important for successful crown gall management.

Therefore, to accomplish this, I characterized 28 virulent *A. tumefaciens* strains collected from the 10 major walnut growing counties of CA for phenotypic (virulence, growth rates, antibiotic resistance, K84 sensitivity) and genotypic diversity.

In California, walnut breeders have been engaged in breeding crown gall-resistant rootstocks to fight against the increasing incidence of crown gall [98]. Current walnut rootstock crown gall-resistance screening uses one or two virulent *A. tumefaciens* strains. Reliance on a limited number of virulent strains in this effort raises several questions. Do these strains represent the diversity of the *A. tumefaciens* population that a given rootstock genotype encounters in the field? Will the rootstock developed be broadly resistant to virulent *A. tumefaciens* strains in walnut orchards? Understanding the genetic diversity of this pathogen will enhance current and future disease resistance breeding efforts.

To answer the above questions, my dissertation research involved the characterization of the diversity of 28 *A. tumefaciens* strains isolated from the top 10 walnut-growing counties in California. One strain CL001 is from a Chilean walnut orchard and the model strain C58 is a reference. Three general areas of *A. tumefaciens* diversity are examined: **First**, I characterized genetic diversity of *A. tumefaciens* strains by performing phylogenetic analyses, calculating whole genome ANI, and pan-genomic analysis in the context of *Agrobacterium* genus. **Second**, I performed phenotypic analysis that evaluated growth rates in rich and poor media, virulence on two hosts, K84 sensitivity, motility, and antibiotic sensitivity. **Third**, I performed comparative genomic analyses to characterize the genetic basis of the observed phenotypic differences, such as the diversity of key genetic loci involved in pathogenesis and competition, e.g., T6SS, T-DNA, opine synthesis, etc.

My dissertation contains five chapters. **Chapter one** is the introduction and literature review. In **Chapter two**, I used 311 strains (282 from NCBI and 29 on this work) for whole genome phylogenetic analyses. A core gene phylogeny was constructed with core gene alignments from the PIRATE pangenome analysis. These results were confirmed with whole genome ANI analysis. In **Chapter three**, the virulence

tests are presented for walnut rootstock and Datura. To limit host genetic variation, the clonal walnut rootstock Vlach was used in these tests. Motility tests, antibiotic resistance tests, growth rate tests, and K84 sensitivity tests of 30 *A. tumefaciens* strains were performed *in vitro*. All results were analyzed using appropriate statistical methods and a Pearson correlation analysis of the five traits was performed to understand their relationships. In **Chapter four**, I evaluated the general genomic characteristics of 30 complete genomes. Then, a comparative genomic analysis is applied to investigate the potential genetic elements underlying the phenotypic differences observed in Chapter 3. Other important genetic traits are also investigated, such as the diversity of T6SS.

The results of these studies will be useful for plant pathologists, plant breeders, and walnut growers. For plant pathologists, they provide comprehensive knowledge and understanding of the diversity of *A. tumefaciens* involved in CG of walnut and its interaction with its host. For walnut rootstock breeders, the information on appropriate strains to be used in crown gall resistance breeding effort is particularly important. For growers, these results provide new perspectives, which will facilitate development of effective crown gall management strategies.

References

1. History. *California Walnuts* https://walnuts.org/about-walnuts/history/.
2. Epstein, L. *et al.* Crown gall can spread between walnut trees in nurseries and reduce future yields. *California Agriculture* **62**, 111–115 (2008).
3. Noncitrus Fruits and Nuts 2018 Summary 06/18/2019. 101 (2018).
4. Rootstock Development - Walnut Research Reports Database. https://ucanr.edu/sites/cawalnut/category/RootDev/?repository=67410&a=173321.
5. Escobar, M. A., Leslie, C. A., McGranahan, G. H. & Dandekar, A. M. Silencing crown gall disease in walnut (*Juglans regia* L.). *Plant Science* **163**, 591–597 (2002).
6. Escobar, M. A. & Dandekar, A. M. *Agrobacterium tumefaciens* as an agent of disease. *Trends in Plant Science* **8**, 380–386 (2003).
7. Escobar, M. A., Civerolo, E. L., Summerfelt, K. R. & Dandekar, A. M. RNAi-mediated oncogene silencing confers resistance to crown gall tumorigenesis. *Proceedings of the National Academy of Sciences* **98**, 13437–13442 (2001).
8. Kim, H. S., Yi, H., Myung, J., Piper, K. R. & Farrand, S. K. Opine-Based Agrobacterium Competitiveness: Dual Expression Control of the Agrocinopine Catabolism (acc) Operon by Agrocinopines and Phosphate Levels. *JB* **190**, 3700–3711 (2008).
9. Kerr, A. Biological Control of Crown Gall: Seed Inoculation. *Journal of Applied Bacteriology* **35**, 493–497 (1972).
10. Garrett, C. M. E. Problems of *Agrobacterium tumefaciens* in planting material and its control1. *EPPO Bulletin* **17**, 263–268 (1987).
11. Ryder, M. H. & Jones, D. A. Biological Control of Crown Gall Using Using *Agrobacterium* Strains K84 and K1026. *Functional Plant Biol.* **18**, 571–579 (1991).

12. Moore, L. W. & Loper, J. E. Fate of *Agrobacterium radiobacter* K84 in the Environmentt. *APPL. ENvIRON. MICROBIOL.* **59**, 9 (1993).
13. Farrand, S. K., Slota, J. E., Shim, J. S. & Kerr, A. Tn5 insertions in the agrocin 84 plasmid: the conjugal nature of pAgK84 and the locations of determinants for transfer and agrocin 84 production. *Plasmid* **13**, 106–117 (1985).
14. Stockwell, V. O. (Horticultural C. R. L., Kawalek, M. D., Moore, L. W. & Loper, J. E. Transfer of pAgK84 from the biocontrol agent *Agrobacterium radiobacter* K84 to *A. tumefaciens* under field conditions. *Phytopathology (USA)* (1996).
15. Penyalver, R. & López, M. M. Cocolonization of the Rhizosphere by PathogenicAgrobacterium Strains and Nonpathogenic Strains K84 and K1026, Used for Crown Gall Biocontrol. *Appl. Environ. Microbiol.* **65**, 1936–1940 (1999).
16. Kim, H. & Farrand, S. K. Characterization of the acc operon from the nopaline-type Ti plasmid pTiC58, which encodes utilization of agrocinopines A and B and susceptibility to agrocin 84. *Journal of bacteriology* **179**, 7559–7572 (1997).
17. Kim, H. S., Yi, H., Myung, J., Piper, K. R. & Farrand, S. K. Opine-Based Agrobacterium Competitiveness: Dual Expression Control of the Agrocinopine Catabolism (acc) Operon by Agrocinopines and Phosphate Levels. *Journal of Bacteriology* **190**, 3700–3711 (2008).
18. Vladimirov, I. A., Matveeva, T. V. & Lutova, L. A. Opine biosynthesis and catabolism genes of *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*. *Russ J Genet* **51**, 121–129 (2015).
19. López, M. M., Gorris, M. T., Temprano, F. J. & Orive, R. J. Results of seven years of biological control of *Agrobacterium tumefaciens* in Spain1. *EPPO Bulletin* **17**, 273–279 (1987).
20. López, M. M., Gorris, M. T., Salcedo, C. I., Montojo, A. M. & Miró, M. Evidence of Biological Control of *Agrobacterium tumefaciens* Strains Sensitive and Resistant to Agrocin 84 by Different

Agrobacterium radiobacter Strains on Stone Fruit Trees. *Applied and Environmental Microbiology* **55**, 741–746 (1989).

21. Farrand, S. K., van Berkum, P. B. & Oger, P. *Agrobacterium* is a definable genus of the family Rhizobiaceae. *International Journal of Systematic and Evolutionary Microbiology,* **53**, 1681–1687 (2003).
22. Gelvin, S. B. *Agrobacterium* Biology: From Basic Science to Biotechnology. (Springer, 2018).
23. Young, J. M., Kuykendall, L. D., Martínez-Romero, E., Kerr, A. & Sawada, H. A revision of Rhizobium Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter, R. rhizogenes, R. rubi, R. undicola* and *R. vitis*. *International Journal of Systematic and Evolutionary Microbiology,* **51**, 89–103 (2001).
24. Holmes, B. & Roberts, P. The Classification, Identification and Nomenclature of *Agrobacteria*. *Journal of Applied Bacteriology* **50**, 443–467 (1981).
25. Ophel, K. & Kerr, *A. Agrobacterium vitis* sp. nov. for Strains of *Agrobacterium* biovar 3 from Grapevines. *International Journal of Systematic and Evolutionary Microbiology,* **40**, 236–241 (1990).
26. Genetello, C. *et al.* Ti plasmids of *Agrobacterium* as conjugative plasmids. *Nature* **265**, 561–563 (1977).
27. Kerr, A. & Panagopoulos, C. G. Biotypes of *Agrobacterium radiobacter* var. *tumefaciens* and their Biological Control. *Journal of Phytopathology* **90**, 172–179 (1977).
28. Young, J. M., Kuykendall, L. D., Martínez-Romero, E., Kerr, A. & Sawada, H. Classification and nomenclature of *Agrobacterium* and *Rhizobium*. *Int J Syst Evol Microbiol* **53**, 1689–1695 (2003).
29. Mousavi, S. A. *et al.* Phylogeny of the *Rhizobium–Allorhizobium–Agrobacterium* clade supports the delineation of Neorhizobium gen. nov. *Systematic and Applied Microbiology* **37**, 208–215 (2014).

30. Kisand, V. & Wikner, J. Limited resolution of 16S rDNA DGGE caused by melting properties and closely related DNA sequences. *Journal of Microbiological Methods* **54**, 183–191 (2003).
31. STACKEBRANDT, E. & GOEBEL, B. M. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic and Evolutionary Microbiology,* **44**, 846–849 (1994).
32. Costechareyre, D. *et al.* Rapid and efficient identification of *Agrobacterium* species by recA allele analysis: *Agrobacterium* recA diversity. *Microb Ecol* **60**, 862–872 (2010).
33. Aujoulat, F. *et al.* Multilocus Sequence-Based Analysis Delineates a Clonal Population of *Agrobacterium (Rhizobium) radiobacter (Agrobacterium tumefaciens)* of Human Origin ▿ . *J Bacteriol* **193**, 2608–2618 (2011).
34. Revised phylogeny of Rhizobiaceae: Proposal of the delineation of Pararhizobium gen. nov., and 13 new species combinations. *Systematic and Applied Microbiology* **38**, 84–90 (2015).
35. Weisberg, A. J. *et al.* Unexpected conservation and global transmission of agrobacterial virulence plasmids. *Science* **368**, (2020).
36. History and current taxonomic status of genus *Agrobacterium* | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S0723202019303418?token=67EABBBB09BF51A46DA419D4F4C3C7544F55003A4545CA8C9909CB7DF02495F6B051A7CBCBA7CD8DF5FDE1092F1C806B doi:10.1016/j.syapm.2019.126046.
37. Tindall et al. - 2010 - Notes on the characterization of prokaryote strain.pdf.
38. Konstantinidis, K. T. & Tiedje, J. M. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**, 2567–2572 (2005).
39. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* **9**, 5114 (2018).

40. Orata, F. D., Meier-Kolthoff, J. P., Sauvageau, D. & Stein, L. Y. Phylogenomic Analysis of the Gammaproteobacterial Methanotrophs (Order Methylococcales) Calls for the Reclassification of Members at the Genus and Species Levels. *Front. Microbiol.* **9**, (2018).
41. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* **36**, 996–1004 (2018).
42. Romalde, J. L., Balboa, S. & Ventosa, A. Editorial: Microbial Taxonomy, Phylogeny and Biodiversity. *Front. Microbiol.* **10**, (2019).
43. Konstantinidis, K. T., Ramette, A. & Tiedje, J. M. The bacterial species definition in the genomic era. *Philos Trans R Soc Lond B Biol Sci* **361**, 1929–1940 (2006).
44. Costechareyre, D., Bertolla, F. & Nesme, X. Homologous Recombination in *Agrobacterium*: Potential Implications for the Genomic Species Concept in Bacteria. *Mol Biol Evol* **26**, 167–176 (2009).
45. Portier, P. *et al.* Identification of Genomic Species in *Agrobacterium* Biovar 1 by AFLP Genomic Markers. *Appl. Environ. Microbiol.* **72**, 7123–7131 (2006).
46. Tettelin, H. *et al.* Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome". *PNAS* **102**, 13950–13955 (2005).
47. Riley, M. A. & Lizotte-Waniewski, M. Population Genomics and the Bacterial Species Concept. in *Horizontal Gene Transfer: Genomes in Flux* (eds. Gogarten, M. B., Gogarten, J. P. & Olendzenski, L. C.) 367–377 (Humana Press, 2009). doi:10.1007/978-1-60327-853-9_21.
48. Rouli, L., Merhej, V., Fournier, P.-E. & Raoult, D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes and New Infections* **7**, 72–85 (2015).
49. Ku, C. *et al.* Endosymbiotic gene transfer from prokaryotic pangenomes: Inherited chimerism in eukaryotes. *PNAS* **112**, 10139–10146 (2015).
50. Tomida, S. *et al.* Pan-Genome and Comparative Genome Analyses of Propionibacterium acnes Reveal Its Genomic Diversity in the Healthy and Diseased Human Skin Microbiome. *mBio* **4**, (2013).

51. Vinuesa, P. & Contreras-Moreira, B. Pangenomic Analysis of the Rhizobiales Using the GET_HOMOLOGUES Software Package. in (2015). doi:10.1002/9781119053095.CH27.
52. González, V. *et al.* Phylogenomic *Rhizobium* Species Are Structured by a Continuum of Diversity and Genomic Clusters. *Front. Microbiol.* **10**, (2019).
53. Sedlazeck, F. J., Lee, H., Darby, C. A. & Schatz, M. C. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nature Reviews Genetics* **19**, 329–346 (2018).
54. Watson, B., Currier, T. C., Gordon, M. P., Chilton, M. D. & Nester, E. W. Plasmid required for virulence of *Agrobacterium tumefaciens*. *Journal of Bacteriology* **123**, 255–264 (1975).
55. Schrammeijer, B. *et al.* Sequence analysis of the vir-region from *Agrobacterium tumefaciens* octopine Ti plasmid pTi15955. *Journal of Experimental Botany* **51**, 1167–1169 (2000).
56. Merritt, P. M., Danhorn, T. & Fuqua, C. Motility and Chemotaxis in *Agrobacterium tumefaciens* Surface Attachment and Biofilm Formation. *Journal of Bacteriology* **189**, 8005–8014 (2007).
57. Shaw, C. H. *et al.* virA and virG are the Ti-plasmid functions required for chemotaxis of *Agrobacterium tumefaciens* towards acetosyringone. *Mol Microbiol* **2**, 413–417 (1988).
58. Tzfira, T. & Citovsky, V. *Agrobacterium: From Biology to Biotechnology*. (Springer Science & Business Media, 2007).
59. Tzfira, T. & Citovsky, V. *Agrobacterium*-mediated genetic transformation of plants: biology and biotechnology. *Current Opinion in Biotechnology* **17**, 147–154 (2006).
60. Hooykaas, Paul. J. J. & Beijersbergen, A. G. M. The Virulence System of *Agrobacterium Tumefaciens*. *Annual Review of Phytopathology* **32**, 157–181 (1994).
61. Hooykaas, P. J. J., Hofker, M., den Dulk-Ras, H. & Schilperoort, R. A. A comparison of virulence determinants in an octopine Ti plasmid, a nopaline Ti plasmid, and an Ri plasmid by complementation analysis of *Agrobacterium tumefaciens* mutants. *Plasmid* **11**, 195–205 (1984).

62. Lacroix, B. & Citovsky, V. Pathways of DNA Transfer to Plants from *Agrobacterium tumefaciens* and Related Bacterial Species. *Annual Review of Phytopathology* **57**, 231–251 (2019).

63. Kado, C. I. Historical account on gaining insights on the mechanism of crown gall tumorigenesis induced by *Agrobacterium tumefaciens*. *Front Microbiol* **5**, (2014).

64. Douglas, C. J., Halperin, W. & Nester, E. W. *Agrobacterium tumefaciens* mutants affected in attachment to plant cells. *Journal of Bacteriology* **152**, 1265–1275 (1982).

65. Douglas, C. J., Staneloni, R. J., Rubin, R. A. & Nester, E. W. Identification and genetic analysis of an *Agrobacterium tumefaciens* chromosomal virulence region. *Journal of Bacteriology* **161**, 850–860 (1985).

66. Transcriptional regulation of the virA and virG genes of *Agrobacterium tumefaciens*. | Journal of Bacteriology. https://jb.asm.org/content/170/9/4047.short.

67. The chvH Locus of *Agrobacterium* Encodes a Homologue of an Elongation Factor Involved in Protein Synthesis | Journal of Bacteriology. https://jb.asm.org/content/183/1/36.long.

68. Schell, J. *et al.* Interactions and DNA transfer between *Agrobacterium tumefaciens*, the Ti-plasmid and the plant host. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **204**, 251–266 (1979).

69. Cangelosi, G. A., Ankenbauer, R. G. & Nester, E. W. Sugars induce the *Agrobacterium* virulence genes through a periplasmic binding protein and a transmembrane signal protein. *PNAS* **87**, 6708–6712 (1990).

70. Kemner, J. M., Liang, X. & Nester, E. W. The *Agrobacterium tumefaciens* virulence gene chvE is part of a putative ABC-type sugar transport operon. *Journal of Bacteriology* **179**, 2452–2458 (1997).

71. Characterization of the mmsAB-araD1 (gguABC) Genes of *Agrobacterium tumefaciens* | Journal of Bacteriology. https://jb.asm.org/content/193/23/6586.long.

72. Bacterial Secretion Systems with an Emphasis on the Chlamydial Type III Secretion System. *Current Issues in Molecular Biology* (2010) doi:10.21775/cimb.012.017.

73. Coburn, B., Sekirov, I. & Finlay, B. B. Type III Secretion Systems and Disease. *Clin Microbiol Rev* **20**, 535–549 (2007).

74. Sgro, G. G. *et al.* Bacteria-Killing Type IV Secretion Systems. *Front. Microbiol.* **10**, (2019).

75. Lin, J.-S., Ma, L.-S. & Lai, E.-M. Systematic Dissection of the *Agrobacterium* Type VI Secretion System Reveals Machinery and Secreted Components for Subcomplex Formation. *PLOS ONE* **8**, e67647 (2013).

76. Mougous, J. D. *et al.* A Virulence Locus of *Pseudomonas aeruginosa* Encodes a Protein Secretion Apparatus. *Science* **312**, 1526–1530 (2006).

77. Leiman, P. G. *et al.* Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *PNAS* **106**, 4154–4159 (2009).

78. Basler, M., Pilhofer, M., Henderson, G. P., Jensen, G. J. & Mekalanos, J. J. Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**, 182–186 (2012).

79. Cianfanelli, F. R., Monlezun, L. & Coulthurst, S. J. Aim, Load, Fire: The Type VI Secretion System, a Bacterial Nanoweapon. *Trends in Microbiology* **24**, 51–62 (2016).

80. Förster, A. *et al.* Coevolution of the ATPase ClpV, the Sheath Proteins TssB and TssC, and the Accessory Protein TagJ/HsiE1 Distinguishes Type VI Secretion Classes. *J. Biol. Chem.* **289**, 33032–33043 (2014).

81. Sgro, G. G. *et al.* Bacteria-Killing Type IV Secretion Systems. *Front. Microbiol.* **10**, (2019).

82. Si, M. *et al.* The Type VI Secretion System Engages a Redox-Regulated Dual-Functional Heme Transporter for Zinc Acquisition. *Cell Reports* **20**, 949–959 (2017).

83. Wu, H.-Y., Chung, P.-C., Shih, H.-W., Wen, S.-R. & Lai, E.-M. Secretome analysis uncovers an Hcp-family protein secreted via a type VI secretion system in *Agrobacterium tumefaciens*. *J Bacteriol* **190**, 2841–2850 (2008).

84. Lin, J.-S., Ma, L.-S. & Lai, E.-M. Systematic Dissection of the Agrobacterium Type VI Secretion System Reveals Machinery and Secreted Components for Subcomplex Formation. *PLOS ONE* **8**, e67647 (2013).

85. Wu, C.-F. *et al.* Plant-Pathogenic *Agrobacterium tumefaciens* Strains Have Diverse Type VI Effector-Immunity Pairs and Vary in In-Planta Competitiveness. *MPMI* **32**, 961–971 (2019).

86. Stockwell, V. O. & Duffy, B. Use of antibiotics in plant agriculture.

87. Aćimović, S. G., Zeng, Q., McGhee, G. C., Sundin, G. W. & Wise, J. C. Control of fire blight (*Erwinia amylovora*) on apple trees with trunk-injected plant resistance inducers and antibiotics and assessment of induction of pathogenesis-related protein genes. *Front Plant Sci* **6**, (2015).

88. Nandakumar, R. *et al. Burkholderia glumae* and *B. gladioli* Cause Bacterial Panicle Blight in Rice in the Southern United States. *Plant Disease* **93**, 896–905 (2009).

89. Distribution and economic importance of fire blight. https://www.cabi.org/isc/abstract/20083014933 (2000).

90. Jones, A. L. & Schnabel, E. L. The development of streptomycin-resistant strains of Erwinia amylovora. in Fire blight: the disease and its causative agent, *Erwinia amylovora.* (ed. Vanneste, J. L.) 235–251 (CABI, 2000). doi:10.1079/9780851992945.0235.

91. McManus, P. S., Stockwell, V. O., Sundin, G. W. & Jones, A. L. Antibiotic use in plant agriculture. *Annu Rev Phytopathol* **40**, 443–465 (2002).

92. Christiano, R. S. C., Reilly, C. C., Miller, W. P. & Scherm, H. Oxytetracycline Dynamics on Peach Leaves in Relation to Temperature, Sunlight, and Simulated Rain. *Plant Disease* **94**, 1213–1218 (2010).

93. Zaman, S. B. *et al.* A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus* **9**, e1403 (2017).

94. Sundin, G. W. & Wang, N. Antibiotic Resistance in Plant-Pathogenic Bacteria. *Annual Review of Phytopathology* **56**, 161–180 (2018).

95. Manyi-Loh, C., Mamphweli, S., Meyer, E. & Okoh, A. Antibiotic Use in Agriculture and Its Consequential Resistance in Environmental Sources: Potential Public Health Implications. *Molecules* **23**, (2018).

96. Merritt, P. M., Danhorn, T. & Fuqua, C. Motility and Chemotaxis in *Agrobacterium tumefaciens* Surface Attachment and Biofilm Formation. *Journal of Bacteriology* **189**, 8005–8014 (2007).

97. Abu-Ashour, J., Joy, D. M., Lee, H., Whiteley, H. R. & Zelin, S. Transport of microorganisms through soil. *Water Air Soil Pollut* **75**, 141–158 (1994).

98. Rootstock Development - Walnut Research Reports Database. https://ucanr.edu/sites/cawalnut/category/RootDev/.

Chapter 2

Phylogenetic, average nucleotide identity, and pangenome analyses facilitate classification and reveal unexpected diversity of 311 *Agrobacterium* strains

Abstract

Agrobacterium is a highly diverse monophyletic genus, whose members cause crown gall or hairy root diseases. To better understand the genetic diversity within the genus, I performed average nucleotide identity (ANI), phylogenetic, and pangenomic analyses of 311 *Agrobacterium* strains, including 29 newly isolated and sequenced strains from California walnut orchards. This collection includes several pathogenic strains of the well-known biovars 1, 2, and 3. ANI analysis revealed much greater diversity than expected among individuals in this group while the pangenome analysis indicated the *Agrobacterium* genus has an open pangenome. Though the sample size was variable, member strains of various ANI groups contained a consistent number of core genes, i.e., 4000-5000 genes. Furthermore, the same four main clades and similar evolutionary relationships were identified using ANI and core gene phylogeny analysis. By combining ANI analysis and pangenome analysis, I provide a robust phylogeny for *Agrobacteria* that was based on whole genomes of 311 strains. This represents one of the most comprehensive phylogenetic analyses of this genus.

Keywords*: Agrobacterium*, pangenome, phylogeny, average nucleotide identity (ANI), core genes, accessory genes, classification

Introduction

The genus *Agrobacterium,* placed in the family Rhizobiaceae by Conn a century ago [1], contains many economically important plant pathogens. For example, *Agrobacterium tumefaciens* and *A. rhizogenes* cause crown gall and hairy root diseases, respectively, whereas *A. vitis* causes crown gall of grapes [2]. The first *Agrobacterium* strain was isolated by Fridiano Cavara in Napoli, Italy from grapevine tumors, which was later confirmed to be *A. tumefaciens* [3]. Since its first discovery*, A. tumefaciens* has been shown to infect most dicotyledonous plants, including many economically important species worldwide, e.g., walnuts (United States), stone fruit (Australia, United States), grapes (Hungary, Bulgaria), and roses (United States) [4]. In 2012, the journal, Molecular Plant Pathology listed

Agrobacterium spp. as one of the top 10 plant pathogenic bacteria based on a poll of molecular plant pathologists. This distinction was due to its role in causing significant crop damage, as well as its role in plant biological and genetic engineering through serving as a model organism to decipher host-pathogen interactions [5].

Classification of Agrobacterium spp. has long been debated 6. The concept of biovars of Agrobacterium was coined as an intrasubspecific category 7, whereas the current usage of biovar is at the species level. The biovar approach grouped Agrobacterium into 3 biovars, biovar 1 (A. tumefaciens), biovar 2 (A. rhizogenes), and biovar 3 (A. vitis). This was based on biochemical tests, serology, pathogenicity, and protein patterns following electrophoresis8 [9]. However, the horizontal transfer of tumor-inducing (Ti) plasmids complicates this classification [10]. Thus, their species names were kept for the pathogenic populations [7]. The plant pathogenic *Agrobacterium* species were *A. larrymoorei, A. rhizogenes, A.rubi, A. tumefaciens*, and *A. vitis* [11]. Young et al. [12] in 2001 proposed to combine the genus *Agrobacterium* into the genus *Rhizobium*. These authors argued that 16S rRNA sequence data did not resolve *Agrobacterium* spp. from *Rhizobium* spp.[13,14,15]. However, Farrand et al. [16] in 2003 suggested that the genus *Agrobacterium* should be retained based on evidence from pulsed-field gel electrophoresis of chromosomes and phenotypic characteristics, such as catabolism of lactose and sucrose [17].

Overall, the classification of *Agrobacterium* spp. has been impacted by limitations of the approaches and criteria used and the lack of consistency in the results from different studies [16]. Advances in DNA sequencing technology now make it possible to examine complete genomes in addition to sequencing multiple alternative single copy housekeeping genes for classification. As sequence data from more housekeeping genes became available, phylogenies built with various housekeeping genes (*recA, mutS, gyrB, glgC*, and *gltD*) have helped clarify *Agrobacterium* spp. classification [18,19]. The high-quality genome sequencing data also makes multilocus phylogeny analysis more powerful.

A Bayesian phylogeny constructed from four concatenated housekeeping genes (*rrs-atpD-recA-rpoB*) indicated *Agrobacterium* is a monophyletic group [20]. A recent multilocus (MLS) phylogeny constructed using 24 conserved genes in Rhizobiaceae species indicated that *Agrobacteria* lineages evolved independently from rhizobia lineages [21]. MLS phylogenetic analyses of Rhizobiaceae have been gradually embraced by the *Agrobacterium* community [6]. A more recent review on *Agrobacterium* classification based on a Neighbor-joining phylogenetic rooted tree using *rrs* (ribosomal rRNA) genes revealed that the genus *Agrobacterium* contained 14 species including *A. tumefaciens*, but *A. vitis* and *A. rhizogenes* were moved into the genus *Allorhizobium* and *Rhizobium*, respectively [22]. However, results from the Bayesian phylogeny, the 24-gene based phylogeny, and the Neighbor-joining phylogeny, the taxonomy of *Agrobacterium* or Rhizobiaceae remains controversial [20,22]. MLS analysis only represents a glimpse into the genome of a given organism. Thus, to obtain an accurate phylogeny, more comprehensive information, such as the complete genome sequence, is required to enhance the resolution of *Agrobacterium* classification.

It has long been accepted that a group of strains showing > 70% DNA-DNA hybridization (DDH) values belong to the same species [23]. This 70% DDH standard has been shown to correspond to ~ 95% average nucleotide identity (ANI) [24,25]. Phylogenomic analysis based on ANI has become a popular tool for the classification of bacteria at both the genus and species levels [26,27]. Phenotypically similar, but genotypically distinct groups of strains have been referred to as genomic species, genomic groups, genomospecies, or genomovars [28,29]. The genomic species concept agrees with species identified using homologous recombination frequencies of closely related *A. tumefaciens* strains [30]. Following this genomic species concept, *A. tumefaciens* has been further divided into 10 genomic species, i.e., G1-G9 and G13 [31]. The species concept based on > 70% DDH value also corresponds to genomic species based on phenotypic and genotypic characteristics (phylogenetic analysis) [30]. Here, I provide evidence to propose that the groups defined by 95% ANI values correspond to a genomic species.

In addition to the species-related concepts described, the Core Genome Hypothesis (CGH) can also be informative to bacterial classification. CGH is a new concept for defining clusters whose members have cluster-specific phenotypic characteristics [32]. The different clusters represent the variation within a population as revealed by pangenome analysis [33]. The pangenome concept first created by Tettelin et al. in 2005, defined the sum of the core genome (all conserved genes), the dispensable genome (accessory genes), and unique genes specific to a given strain [34,35]. Pangenome analysis has become increasingly popular as a method to reveal and identify important genetic elements involved in pathogenesis, antibiotic resistance, etc. [34,36]. Pangenome analysis has been applied to study the evolution of host-microbe interactions and niche adaptation for *Rhizobiales* [37]. Also, pangenome analysis has been used to facilitate classification. For example, the pangenome analysis of the order Rhizobiales successfully distinguished the genera *Agrobacterium* and *Rhizobium* from other genera in this family [37]. The gene presence/absence profile of the *Rhizobium* pangenome analysis confirmed the phylogenomic pattern of divergence of species in the genus *Rhizobium* [38]. Therefore, pangenomics can be used to examine the genomic content of ANI groups. This will facilitate answering the following questions: Are there unique genes that define the boundary of an ANI group? Will the number of core genes decrease as the number of strains examined increases in an ANI group? Applying pangenome and ANI analyses to understand classification and the relationships of *Agrobacterium* spp. has not been performed.

To fill the *Agrobacterium* classification gaps mentioned above, it is important to create a whole genome-based phylogeny, to calculate the whole genome ANI values, and to perform pangenome analysis of *Agrobacterium* spp. Here, I take advantage of cost effective sequencing technologies, publicly accessible whole genome sequences, and readily available bioinformatics tools [39] to examine the classification of 311 *Agrobacterium* strains. I first performed whole-genome ANI analyses, which placed 311 strains into 35 ANI groups. A phylogeny was then created with the ANI values to illustrate their

relationships. Next, a pangenome analysis of the 311 strains was performed at the ANI-level and clade-level. A phylogeny was constructed from core-gene alignments to further reveal their evolutionary relationship. The classification results from ANI, pangenome, and phylogenetic analyses were congruent. The results of this study extend our current understanding of *Agrobacterium* spp. diversity and classification, which will facilitate our fundamental understanding of *Agrobacterium* biology and ecology [6].

Methods

Agrobacterium genome retrieval from database and sequencing of new strains

Two sources of *Agrobacterium* sequences were used in this study. The genomes of 296 *Agrobacterium* strains were downloaded from the NCBI RefSeq database on July 20th, 2020 (Table S1). After retrieving the sequences, the genomes were quality-checked with CheckM [40]. I excluded six low-quality genomes and eight duplicate depositions of higher quality genomes. The remaining 282 genome sequences had a completeness greater than 99% (threshold >90%) and a contamination level of less than 5% (threshold < 5%) (Table S2). Of the 282 genomes, 28 were complete genomes and 254 were draft genomes. Four *Sinorhizobium meliloti* (USDA1157, SM11, Rm41, and KH46) strains with high-quality complete genome sequences were selected to form an outgroup for phylogenetic analyses.

The second source of *Agrobacterium* sequences consisted of 29 *Agrobacterium* strains that were isolated, sequenced, assembled, and annotated in the present study. Among these 29 strains, 28 were isolated from crown gall tissue, wood stem tissue, and soil from walnut orchards located in the top 10 walnut growing counties in California from 2008 to 2017, including Tehama, Glenn, Butte, Sutter, Yuba, Yolo, San Joaquin, Stanislaus, Kings, and Tulare. The other *A. tumefaciens* strain CL001 was isolated from a walnut orchard near Santiago Chile in 2017. Genomic DNA of all 29 strains was extracted using MasterPure™ genome DNA Extraction Kit (Epicentre, Middleton, WI). Genomic DNA QC was performed using standard gel electrophoresis for quality, a Qubit Fluorometer for quantity, and a Nanodrop

Spectrophotometer (Nucleic Acid 260/280 ratio) for purity. High quality genomic DNA were then submitted for both Illumina NGS MiSeq and third-generation Nanopore sequencing. Nanopore sequencing was performed at the Genomics Core – Research Technology Support Facility at Michigan State University, MI, and Illumina MiSeq (2x250 PE reads) sequencing was performed at the UC Davis DNA Technologies Core facility, Davis, CA. Hybrid assembly was performed with Unicycler [41], and assembly quality was evaluated with CheckM [40]. Genome assemblies were annotated with the Prokaryotic Genome Annotation Pipeline (PGAP) from NCBI [42,43]. The complete genome of *A. tumefaciens* strain Sta001 (GenBank accession number NNAQ00000000) was submitted to NCBI in 2019 to replace the draft genome submitted in 2017 [44]. The high-quality whole genome sequences of the 29 *A. tumefaciens* strains were obtained for further data analyses. In total, 57 complete genomes and 254 draft genomes of 311 *Agrobacterium* strains and four *S. meliloti* genomes were included in our analysis, i.e., 315 total sequences.

Whole-genome average nucleotide identity (ANI) analyses

FastANI was applied to all 315 whole-genome sequences to calculate pairwise average nucleotide identity (ANI) values [25]. The strains were classified into different ANI groups based on the criterion of ANI value ≥ 95% [29]. Including the complete genome sequences of the four *S. meliloti* strains (USDA1157, KH46, SM11, and Rm41), representing strains of the outgroup, also allowing the ANI pairwise identity values to be converted to a PHYLIP distance matrix. A Neighbor-joining tree was created from this distance matrix of 315 strains using the BIONJ algorithm [45]. An ANI heatmap was generated using custom R script. A table was created showing each ANI group and corresponding NCBI taxonomy of strains in these groups. For example, strain C58 is classified in group ANI8 and is a strain of *A. fabrum* according to NCBI taxonomy and belongs to genomic species 8 (G8) (Table S3). In this chapter, I used the NCBI taxonomy to understand the classification of *Agrobacterium* spp. and compare ANI groups.

Pangenome construction of 311 *Agrobacterium* spp.

All 311 *Agrobacterium* genomes were annotated using Prokka [46]. The annotated genomes were then used as input for the pangenome analysis pipeline Pangenome Iterative Refinement and Threshold Evaluation (PIRATE) [47] following the GitHub tutorial (https://github.com/SionBayliss/PIRATE). The PIRATE pipeline applies variable threshold to identify paralogous gene families that are homologous to those involved in speciation events. All genes were aligned using mafft 7.310 built into PIRATE, and core alignments were generated with "-a" flag in the command. PIRATE plot summaries using R 3.4.1 with "--rplots" flag in the command were also generated. A maximum likelihood (ML) Core phylogeny was constructed using core alignments by running FastTree using the GTR+CAT model and default settings [48].

The output file called PIRATE.gene_families.ordered.tsv was the main file used in this study and was converted to contain original locus_tag/IDs using subsample_outputs.pl as explained in the tutorial. This converted version, i.e., named as the final version, was further converted to Roary format using PIRATE_to_Roary.pl to make the data compatible with the Roary pipeline [49]. The Roary version of PIRATE_gene_presence_and_absence csv file was used for pangenome statistical analyses, and custom R and Python scripts were applied. Pangenome visualization was executed using roary_plots.py following Roary GithHub tutorial (https://github.com/sanger-pathogens/Roary). Roary version 3.11.2 was used for pangenome statistical analyses and visualization. Custom R and Python scripts were applied for data analyses.

I determined the pangenome composition of each clade and each ANI group. The members of each ANI group and each clade were obtained from the tips/leaves of the ANI-based and core genes phylogenies. I then used these data as an index to subset the Roary version of PIRATE_gene_presence_and_absence_csv file. The gene_presence-absence csv file for each ANI and clade were then created. These csv files were further used to calculate the pangenome, core genes, accessory genes, core/pangenome ratio, and core/accessory ratio. The core/pangenome ratio was used

to define species boundary of *Klebsiella* genus [50]. Thus, I calculated the core/pangenome to evaluate species in the genus *Agrobacterium*. Because the accessory genes represent the difference of one strain from the other, I also calculated the core/accessory ratio to indicate strain differences in the same ANI group. A small core/acc ratio indicates a greater strain difference due to a greater number of accessory genes. The corresponding R codes for these analyses were specified in the R markdown file called coreGnechp2.Rmd.

Functional annotation of genomes

Our ANI-based and core genes phylogenies analyses revealed four main clades of 311 strains. To examine the differences in functional genes between the *Agrobacterium* strains, the gene presence and absence profile of each clade was extracted from the final version of the gene presence/absence file. An index file for each clade, with each gene family having one representative locus-tag, was then created. This index file was used as a reference to extract coding sequences (CDSs) for each clade from the Prokka output and had the clade CDS annotate by EggNog-mapper [51,52].

R packages and Server platform

All bioinformatic processing was done using the USDA SCINet Ceres Cluster. R packages used in this study included disparity [53,54,55], ggplot2 [56], ggtree [57], treeio [58], tidyverse [59], tidyr [60], tidyselect [61], SparseM [62], scales [63].

Results

Classification of *Agrobacterium* species into ANI groups

I successfully classified the 311 strains into 35 ANI groups (ANI1-ANI35) (Fig. 2.1; Table 2.1). Based on the ANI Neighbor-joining phylogeny, these ANI groups fell into four major clusters/clades: *A. tumefaciens* strains (cluster 1), *A. rhizogenes* strains (cluster 3), *A. vitis* strains (cluster 4), and a mixture of other species and unclassified *Agrobacterium* spp. (cluster 2) based on NCBI taxonomy (Fig. 2.2).

Comparing the ANI values of each cluster, I found that cluster 1 ANI values ranged from 81.6 to 100%; cluster 2 ANI values ranged from 80.6 to 100%; cluster 3 ANI values ranged from 84 to 100 %; and cluster 4 ANI ranged from 92.80 to 100% (Fig. 2.1)

I compared members in the 10 genomic species of the *Agrobacterium tumefaciens* (G1-G9, G13) complex with members of the 35 ANI groups. For this comparison, the two most recently published phylogenetic trees were selected as reference trees: 1) a maximum likelihood phylogeny built from 400 single house-keeping genes [6] and 2) a Neighbor-joining phylogeny constructed from *rrs* gene sequences [21]. The members of G1 all grouped into ANI1. However, 11 strains in ANI1 had unknown genomic species assignments. Given the overall 95% ANI pairwise identity of strains in this group, I assign these 11 strains to G1. These strains were But001, Sta004, CL001, Yub002, CNPSo675, SBV_303932, SBV_38, RS6, B41, ATCC4720, and ICMP4364. The first four strains listed were from this study and the rest were from NCBI. Following this above comparison, I successfully assigned 55 unclassified strains from eight ANI groups into corresponding genomic species marked red (Table S3) based on the 95% ANI values. I also found that the 10 genomic species (G1-G9, G13) corresponded to ANI1 through ANI10, respectively. ANI14, ANI15, ANI21, ANI25, and ANI31, all of which were members of *A. tumefaciens,* were closely positioned with ANI1-ANI10 and formed cluster 1 in the ANI-based phylogeny (Fig. 2.2). I also suggest that two strains (NCPPB2659 and GBBC3284) that were originally grouped in *A. rhizogenes* should be considered as *A. tumefaciens* based on the ANI-phylogeny analysis.

The ANI11, ANI13, ANI16, ANI17, ANI22, ANI29, and ANI30 groups formed a cluster (cluster 2), whereas ANI12, ANI24, ANI32, ANI33, ANI34, and ANI35 were grouped in cluster 3. ANI11 corresponded to *A. larrymoorei* based on ANI analysis. ANI13 corresponded to *A. rosae*, and this included two strains DSM25558 and DSM25559. Members of ANI16 consisted of strains of *A. fabrum*, *A. tumefaciens*, and *A. rhizogenes,* as well as unclassified strains. ANI17 corresponded to *A. rubi* and ANI22 corresponded to *A. bohemicum*. Members of ANI groups in cluster 3 were strains of *A. rhizogenes*. ANI18-20 and ANI28

belonged to cluster 4. Most strains in cluster 4 were *A. vitis*. Among the 35 ANI groups, 11 contained only one strain (Table 2.1).

Core genes phylogeny of *Agrobacterium* species

To understand the evolutionary relationship of the 311 strains, a core gene phylogeny was constructed using maximum likelihood (ML) phylogenetic inference from core gene alignments rooting with the common ancestor of RAC06, MA01, and AOL15 (Fig. 2.3), which were the closest group to the outgroup in the ANI phylogeny. Similar to the ANI-based tree, this phylogeny revealed four main clades (C1-C4). Interestingly, members of these four clades corresponded to members of the four clusters revealed using the ANI heatmap (Fig. 2.1). Clade 1 is comprised of 15 ANI groups: ANI1-10, ANI14, ANI15, ANI21, ANI25, ANI31. Members in this clade were strains of *A. tumefaciens* (Table 2.1). Clade 2 consisted of 7 ANI groups: ANI11 (*A. larrymoorei*); ANI13 (*A. rosae*); ANI16; ANI17 (*A.rubi*); ANI22 (*A. bohemicum*) ANI29, and ANI30. Clade 3 contained 6 ANI groups: ANI 12, ANI 24, ANI 32-35. Members in Clade 3 were primarily *A. rhizogenes*. Clade 4 included 4 ANI groups: ANI 18-20, ANI28, all of which were *A. vitis*. Four strains, RAC06 and MA01 (ANI23), A0L15 (ANI26), and a222 (ANI27), were not part of the four main clades.

Diversity of *Agrobacterium* strains from walnut orchards in CA

The core gene phylogeny showed that the 29 strains isolated from California walnut orchards were assigned to eight ANI groups. The majority of strains (17) belonged to ANI4. The others were in ANI1 (4 strains), ANI7 (3 strains), and five ANI groups (ANI2, ANI8, ANI10, ANI15, and ANI25) each with one strain. The members in ANI4 were the most common of the *A. tumefaciens* strains detected in CA walnut orchards, including from 8 counties: Stanislaus, San Joaquin, Butte, Tehama, Tulare, Sutter, Glenn, and Yolo. Strains recovered from walnut orchards in Kings County had the greatest diversity with representatives from ANI7 (Kin003), ANI8 (Kin001), ANI15 (Kin002). Interestingly, strain CL001, isolated from a walnut orchard in Chile, was closely clustered with strain Sta004 from Stanislaus County in CA.

Topology comparison between ANI phylogeny, core phylogeny, and MLS phylogeny

The ANI Neighbor-joining phylogeny of 311 *Agrobacterium* genomes was rooted to the outgroup, which is the common ancestor of the four *S. meliloti* strains (USDA1157, KH46, SM11, and Rm41) (Fig. 2.2). The core gene phylogeny of the 311 *Agrobacterium* genomes was rooted at the common ancestor of members in group ANI23 and ANI27 (Fig. 2.3). These two ANI groups were most closely clustered with the outgroup in the ANI phylogeny (Fig. 2.2). The MLS phylogeny is rooted with the same outgroup as the ANI-based phylogeny. Topology comparison showed that the topology of ANI phylogeny was congruent with the core gene phylogeny in terms of the four clades/clusters (C1, C2, C3, and C4), respectively (Fig 2.5). However, the ANI-based phylogeny and MLS phylogeny exhibited different tree topologies (Fig. 2.6). C1 and C2 shared the same topology, whereas C3 and C4 had difference in the ANI-based and MLS phylogenies. Furthermore, within the same clade, the topology of ANI groups of the three phylogenetic trees were all variable (Fig. 2.5; Fig. 2.6).

Pangenome analysis of *Agrobacterium* at species/ANI level

The Roary pangenome pipeline designates core genes, soft-core, shell, and cloud genes. Core genes are those found in greater than 99% of the genomes examined; soft-core genes are found in 95% to 98% of the genomes; shell genes are found in 15% to 94% of the genomes; cloud genes are found in less than 14% of the genomes. The PIRATE pangenome is the sum of the core gene family, which corresponds to the sum of core genes and soft-core genes in the Roary pipeline; and the accessory gene family, which is the sum of shell and cloud genes in the Roary pipeline.

The pangenome of the 311 *A. tumefaciens* strains examined using PIRATE pipeline consisted of 52,233 gene families. Only 1,967 (3.8%) gene families were classified as core and soft-core genes and present in > 95% of the genomes examined. Furthermore, 210 of the 1,967 core genes were soft-core genes. These soft-core genes were present in more than 295 strains, but in less than 307 strains (Fig. 2.7). Besides the core genes, the remaining 50,266 gene families (96.2%) constituted of accessory genes.

These accessory genes consisted of 7,070 shell genes and 43,196 cloud genes. Among the accessory genes, 42,290 were present in less than 31 genomes (Fig. 2.7; Table 2.2).

To further examine the pangenome at the species level, I included the 22 ANI groups that contained 2 or more strains. ANI22 and ANI24 were excluded because they contained less than 50 accessory genes for each group indicating I am not confident to include these ANI groups for further data analysis. On average, the 22 ANI groups displayed diverse pangenomes, ranging from ~4,000 to ~15,000 genes.

The number of core genes (~4,800-5,000) within 22 ANI groups was consistent compared with their pangenome (~5,000-15,000) (Fig. 2.8). The ANI group with the largest pangenome had a greater number of accessory genes, because the pangenome is the sum of core and accessory genes (Fig. 2.8). The core/pangenome ratio of each strain in one ANI group varied. For example, in ANI7, this ratio ranged from 0.65 to 0.82, whereas in ANI4, it ranged from 0.71 to 0.88 (Tables in the Github: https://github.com/limin321/dissertation_codes/tree/main/ANIgroup_subset/strainCor_accRatio).

The ANI group with the highest core/acc ratio variance contained strains that exhibited lower variability. The ANI group with the lowest variance contained more highly variable strains (Fig. 2.9; Table S4). I also calculated core/accessory gene (core/acc) ratio for members in each ANI group to evaluate strain variation within each ANI group. Variance in the core/acc ratio in most ANI groups was quite low with the variance value < 0.5, whereas ANI4 and ANI9 had a variance of 1.1 and 8.0, respectively (Table S4). ANI16, ANI17, ANI18, and ANI20 had a similar interquartile range (IQR) for the core/acc ratio. ANI7, ANI8, and ANI12 had similar IQR of the core/acc ratio. ANI1, ANI4, and ANI13 had similar IQR. ANI9 had the widest IQR of the core/acc ratio ranging from 2.3 to 8.6 (Fig. 2.9).

Due to the uneven number of strains in each ANI group, I also ran ANI-level pangenome comparisons. The 22 ANI groups were divided into two subsets, subset 1 containing ANI groups with more than 20 strains (ANI1, 4, 7, 12, 20), and subset2 including ANI8, 9, 13, 16, 17, 18, 19, each

containing 6 to 19 strains. In subset 1 (Fig. 2.10), most ANI groups had a similar pangenome with ~5,000 genes. However, ANI12 had the biggest pangenome with ~6,400 genes. Most ANI groups contained approximately 3,800 core gene, whereas ANI12 had around 5,000 genes. Interestingly, the number of accessory genes of each strain was similar in each ANI group, i.e., ~1,500 genes (Fig. 2.10).

In subset 2, the pangenome of most strains was approximately ~ 5,000 genes for most ANI groups, whereas the pangenome of each strain in ANI16 possessed ~4500 genes. Strains in ANI9 had a much larger variance than strains in other ANI groups, in terms of the pangenome. For accessory genes, most strains contained nearly 1,000 genes in most ANI groups, while strains in the ANI9 had more variance than strains in other ANI groups (Fig. 2.11). The number of core genes of each strain in both subsets (Fig. 2.10; Fig. 2.11) were similar containing ~3,800 genes. In subset1, the pangenome on average was greater than 5000, whereas it was < 5000 for subset2. Strains in subset1 had > 1200 accessory genes, whereas those in subset2 had ~1000.

To examine the unique core genes of each ANI group, I compared the core genes of each ANI group with all genes in other ANI groups. I found that ANI4 (42 strains) and ANI7 (24 strains) had 0 unique core genes (Table S5). The 3 strains in ANI11 had the greatest number of 129 unique core genes. ANI12 (76 strains) had 96 unique core genes. ANI14 had 93 unique core genes, though it only included two strains. Core genes of each ANI group in the *A. tumefaciens* clade were also compared with genes of other ANI groups in the *A. tumefaciens* complex. Similarly, ANI4 and ANI7 had no unique core genes (Table S6). All other ANI groups in the *A. tumefaciens* clade contained more unique core genes when compared with ANI groups from the *Agrobacterium* genus.

Core genes of each ANI group were compared with those in all other ANI groups. The results showed that each ANI group had ANI-specific core genes and the number of these genes was different among ANI groups (Table S7). When comparing core genes of one ANI groups with ANI groups from the

same clade, i.e., *A. tumefaciens,* the number of ANI-specific core genes in that ANI group increased (Table S8). I identified 7 ANI-specific core genes in ANI4 and 2 ANI-specific core genes in ANI7.

Pangenome analysis of *Agrobacterium* spp. on clades level

The previous ANI-based and core gene phylogenetic analyses successfully grouped 311 *Agrobacterium* strains into four main clades (Fig. 2.2; Fig. 2.3). I examined the pangenome of each clade (Table S9). The pangenome for clades 1, 2, 3 and 4 were 29540, 13789, 19607, and 12275 genes, respectively. The number of core genes for clades 1, 2, 3, and 4 was 2602, 2481, 3486, and 3532, and the number of accessory genes was 26938, 11308, 16121, and 8743, respectively. To study the gene presence/absence signature of each clade, I focused only on core genes and compared them with one another. Interestingly, 279, 154, 1051, and 1137 clade-specific core genes for clade 1, 2, 3, and 4, respectively, were found. More details of gene family components could be found in Table S10.

Clusters of Orthologous Genes (COGs) analysis of clades

To understand what genes may define each clade, I performed COGs analysis of clade-specific core genes. Clade-specific core genes of each clade were functionally annotated using EggNog-mapper software [64,52]. Interestingly, COGs greater than 2% were function unknown (S), transcription (K), carbohydrate transport and metabolism (G), inorganic ion transport and metabolism (P), cell wall/membrane/envelope biogenesis (M), signal transduction mechanism (T), amino acid transport and metabolism (E), energy production and conversion (C), lipid transportation and metabolism and secondary metabolites biosynthesis, transport, and catabolism (IQ) in clade 1 (Table S11). In clade 2, COGs greater than 2% included function unknow (S), signal transduction mechanism (T), cell wall/membrane/envelope biogenesis (M), inorganic ion transport and metabolism (P), amino acid transport and metabolism (E), carbohydrate transport and metabolism (G), transcription (K), cell motility (N), second metabolites biosynthesis, transport, and catabolism (Q), lipid transportation and metabolism (I), coenzyme transport and metabolism (H), replication, recombination and repair (L) (Table S12).

Similarly, in clade 3, function unknow (S), transcription (K), carbohydrate transport and metabolism (G), amino acid transport and metabolism (E), cell wall/membrane/envelope biogenesis (M), energy production and conversion (C), inorganic ion transport and metabolism (P), signal transduction mechanism (T), lipid transportation and metabolism (I), and EGP were identified (Table S13) and in clade 4, function unknow (S), transcription (K), amino acid transport and metabolism (E), inorganic ion transport and metabolism (P), carbohydrate transport and metabolism (G), signal transduction mechanism (T), cell wall/membrane/envelope biogenesis (M), energy production and conversion (C), lipid transportation and metabolism (I), second metabolites biosynthesis, transport, and catabolism (Q), post-translational modification, protein turnover, and chaperones (O), and defense mechanism (V) were identified (Table S14). All four clades contained ~25% unknown function COGs represented by S.

All four clades contained COGs involved in amino acid metabolism and transport (E), carbohydrate metabolism and transport (G); lipid metabolism (I); transcript (K); cell wall/membrane/envelop biogenesis (M); inorganic ion transport and metabolism (P); signal transduction (T). However, the percent of some of the above COGs were significantly different. G in clades 2 and 4 was ~7%, whereas in clades 1 and 3 alone, it was > 10%. K in clade 2 only had 3.2 %, whereas others had more than 10%. M in clade 1 and 2 with ~8.5% was higher than those of clade 3 and 4 with ~6.3%. P in clade 3 had the smallest amount (~4.2%), whereas it was > 7% in others. T was the most variable COGs with 7.7% in clade1, 11.3% in clade2, 3.7% in clade 3, and 6.9% in clade 4. N (cell motility) and L (replication and repair) were mostly found in clade 2. O (post-translational modification, protein turnover, chaperone functions) was mostly found in clade 4.

Clusters of Orthologous Genes (COGs) analysis of ANIs

To identify what genes may be common to ANI groups in each clade, I assigned unique core genes of ANI groups containing more than 20 strains to COGs. Using this criterion, ANI1, ANI4, ANI7 in clade 1 were chosen for this analysis. ANI1 had 56 unique core genes, which were clustered into 10 main

COGs. They were 24.4 % amino acid metabolism and transport (E); 17.8% transcription (K); 13.3 % cell motility and signal transduction (NT); 13.3 % unknown function (S); 8.9% energy production and conversion (C); 6.7% signal transduction (T); 4.4% inorganic ion transport and metabolism (P); 4.4% amino acid metabolism and transport and signal transduction (ET); 4.4% lipid metabolism and secondary structure (IQ); 2.2% lipid metabolism (I) (Table S15). However, ANI4 had 7 unique core genes, some of which were functionally annotated relating to chromate resistance, N-methylhydantoinase A acetone carboxylase, beta subunit, DinB family, BrnA anti-toxin of type II toxin-anti-toxin system (Table S7). ANI7 had 2 unique cores genes, both of which were hypothetical proteins of unknown function (Table S7).

Discussion

Phylogeny analyses of *Agrobacterium* strains

The *Agrobacterium* genus contains various economically important pathogens, such as the *A. tumefaciens* complex and *A. rhizogenes*, which cause crown gall and hairy root on host plants respectively, and *A. vitis* that induces crown gall on grapes worldwide [65]. Classification of strains in this genus has long been a controversial subject among the *Agrobacterium* research community [22]. In this study, I performed whole genome ANI, core gene phylogeny, 6 single house-keeping genes (MLS) phylogeny analysis and pangenome analyses of 311 *Agrobacterium* genomes. In doing so, I found that 4 major clades/clusters (C1, C2, C3, and C4) were identified by these approaches. The evolutionary relationship of C1, C2, C3, and C4 revealed by ANI and core phylogeny was in agreement with that found with the phylogeny for the 24 housekeeping genes [21]. However, this relationship was incongruent with the 6-gene MLS phylogeny, indicating that whole-genome based phylogeny provides a more comprehensive view into the evolutionary relationship of bacterial groups.

The four clades revealed in the whole-genome phylogeny were further subdivided into 15, 7, 6, and 4 ANI groups, respectively (Fig. 2.2). Among the 15 ANI groups in C1, strains in ANI1-10 correspond to the 10 described genomic species, G1-G9 and G13 [31]. This result is also consistent with the 24-gene

based phylogeny with the exception of G7, strains of which were placed in two different branches (Fig. 1) [21]. Members in ANI groups of the *A. tumefaciens* clade were the same as members as found in the corresponding genomic species, which was also confirmed with the 400 single-copy genes phylogeny [6]. Thus, I conclude that the 15 ANI groups in C1 can be considered 15 genomic species. The ANI analysis also classified C4 (*A. vitis*) into 4 ANI groups. Genomic comparisons of the *A. vitis* strains showed that they were heterogenous and belonged to at least 3 genomic species [6]. A recent study revealed that *Allorizobium vitis* (*A. vitis*) is a species complex consisting of 4 genomic species [66]. Together, these data support the concept that an ANI groups correspond to a genomic species. This also provides evidence of genetic diversity in *A. vitis*. Following this concept, C2 contains at least 7 genomic species and C3 (*A. rhizogenes*) contains at least 6 genomic species. In fact, genomic species have been given a species name [67]. For example, C58 of *A. tumefaciens* G8 was identified as *A. fabrum* [67]. Therefore, I propose here that each ANI group could be assigned a species name, and each clade be given a genus name for clear classification.

The topology analysis of the ANI-based phylogeny, core genes phylogeny, and MLS phylogeny showed that the ANI-based and core phylogeny were consistent with each other at the clade level and inconsistent with that of the 6-gene MLS phylogeny. Furthermore, the topology of these three phylogenies were variable at the ANI level, whereas the members of each ANI group were the same as members of the corresponding group in other phylogenies. These results indicate that whole genome-based phylogeny is essential to more clearly resolve *Agrobacterium* strain classifications. As more high-quality whole genome sequences become available, more ANI/genomic species continue to provide insight into the true diversity of *Agrobacterium* strains.

Pangenome analysis of *Agrobacterium* strains

Discontinuous variation is a central idea for conventional taxonomy, indicating there is no transition between different species [68]. For example, the core/pan-genome ratio was used to reclassify

six strains of *Klebsiella* spp. whose core/pan-genome ratios were greater than 94% and the ratio break between two species was greater than 10% [50]. However, this core/pangenome ratio is more applicable for bacteria with a closed pangenome. It is more challenging to apply the core/pangenome ratio to define species with an open pangenome [69]. In my study, pangenome analysis of 311 *Agrobacterium* strains reveals that *Agrobacterium* spp. have an open pangenome. The ~96% of the pangenome consists of accessory genes indicating there is great variation within *Agrobacterium* spp. Members in the ANI groups exhibited core/pan-genome ratios ranging from 0.65 to 0.99 (Table 2.3). This broad ratio range also indicated great diversity among the members in each ANI group. This further implies that core/pangenome ratio is not an ideal parameter to define species in the *Agrobacterium* genus in contrast to the genus *Klebsiella* [50].

The consistent number of core genes for strains in each ANI group (~4800) compared with the wide range of accessory genes (~2000 - ~10000) also supports the conclusion that *Agrobacterium* spp. have an open pangenome (Fig. 2.10). The gene presence/absence profile of 311 strains from pangenome analysis confirms the 4 clades/clusters and 35 ANI groups generated by ANI and whole genome phylogeny analysis (Fig. 2.13.).

The divergent strain core/acc ratio of the strains in one ANI group indicated individuals within a species are variable. That is, one individual clearly cannot represent a single species. Comparison of core genes among different ANI groups demonstrates that the boundary for the ANI group is not defined by certain unique genes. Instead, they should be defined by the combination of genes from the whole gene repertoire (or pangenome).

Pangenome analysis at the clade level also revealed that the boundary of a given clade is defined by the combination of genes from the pangenome. The close phylogenetic relationship between C1 and C2 may explain why C1 and C2 share a similar number of core genes. C1 contains 149 strains and has fewer core genes than other clades containing less than 85 strains. This indicates the number of core

genes decrease as more members are added to the clade. The core/acc ratio for C1 and C4 is 0.10 and 0.4, respectively, indicating C1 consists of individuals with greater heterogeneity. This is in agreement with the 95% ANI value threshold revealing 15 ANI groups in C1 and 4 ANI groups in C4.

In summary, this study illustrated an unexpected diversity among the *Agrobacterium* strains examined. ANI analysis, core-gene phylogeny, and pangenome analysis are in concurrence and support the reclassification of *Agrobacterium* spp.

Figures and Tables

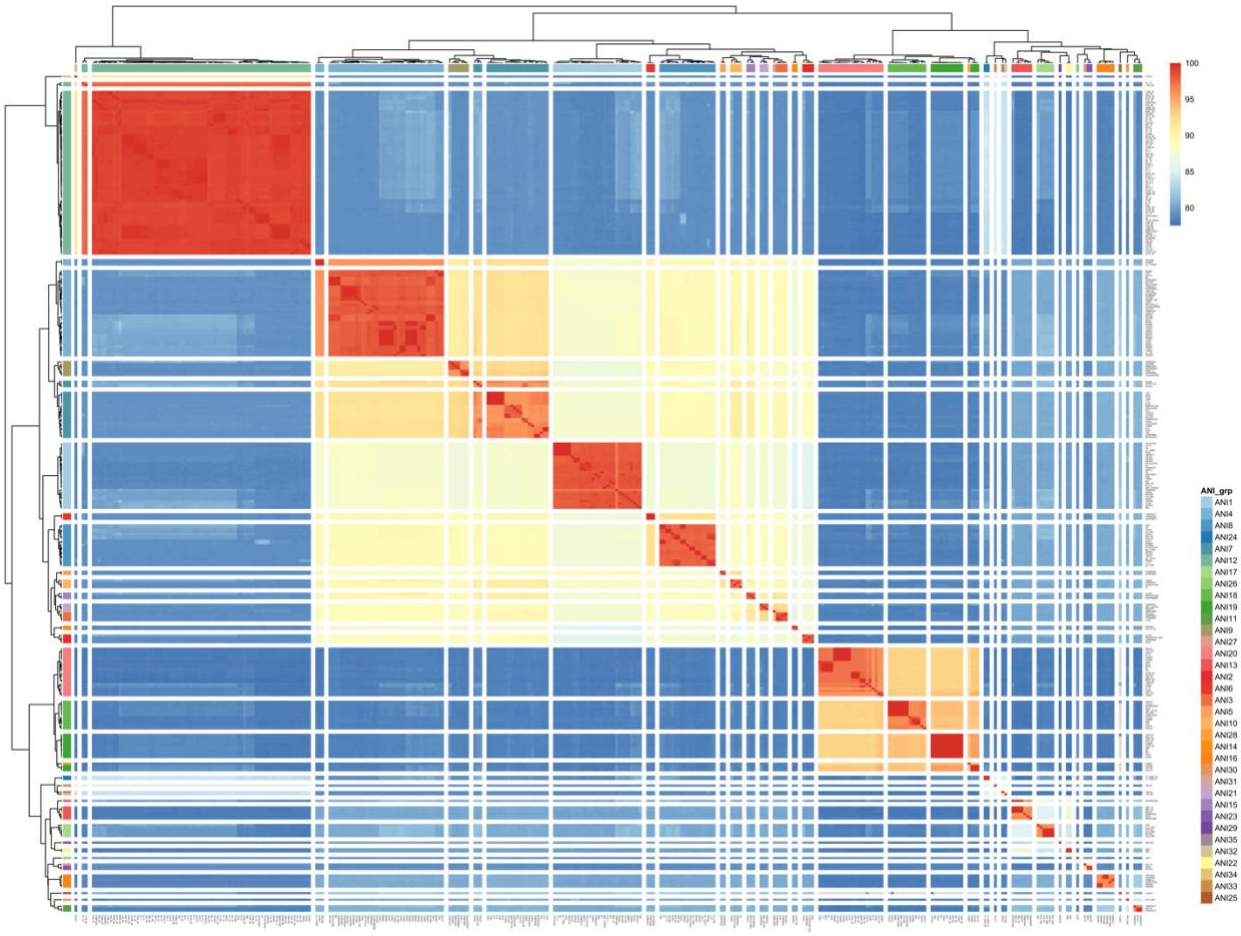


Fig. 2.1. Heat map of 311 *Agrobacterium* strains clustered into ANI groups. The 35 ANI groups were generated based on ≥ 95% pairwise identity between ANI values. Top and left dendrograms are based on complete-linkage hierarchical clustering. ANI groups designations are mapped via color code. The right and bottom labels are the 311 Agrobacterium strains.

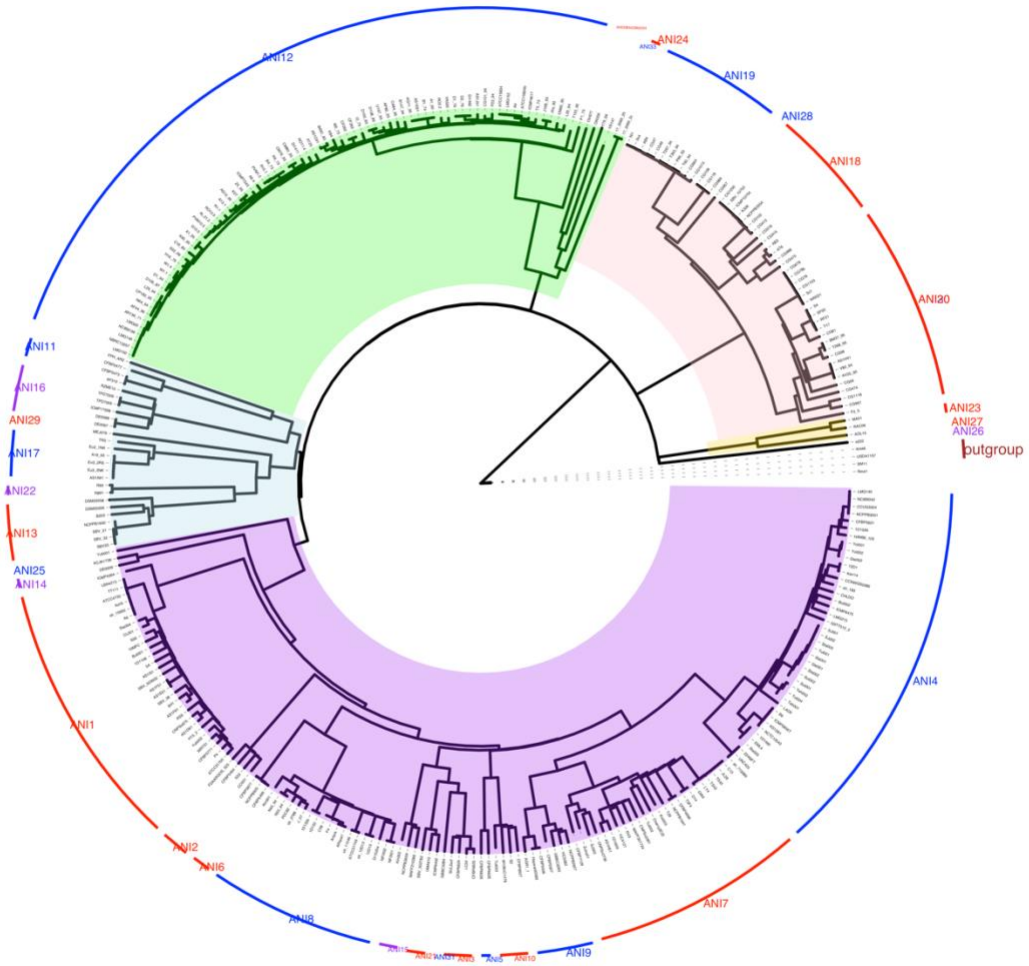


Fig. 2.2. The ANI-based phylogeny of 311 *Agrobacterium* genomes. The ANI phylogeny is a Neighbor-joining tree constructed from a distance matrix of ANI pairwise values. *Sinorhizobium meliloti* isolates were used as an outgroup. The four main clusters are highlighted in color. Purple: cluster 1 (C1; *A. tumefaciens*); Light blue: cluster 2 (C2; mixture of named and unclassified *Agrobacterium*); Green: cluster 3 (C3; *A. rhizogenes*); Pink: cluster 4 (C4; *A. vitis*). The outside ring shows the 35 ANI groups.

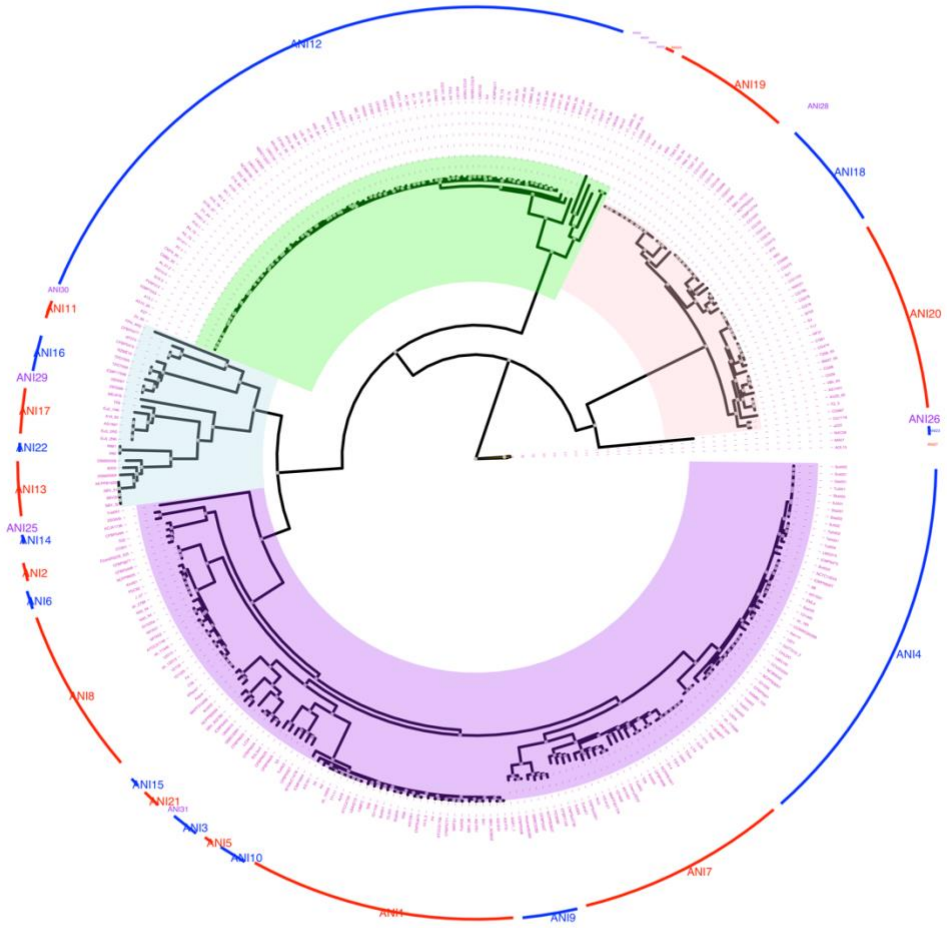


Fig. 2.3. The core genes phylogeny of 311 *Agrobacterium* strains. This maximum likelihood phylogeny was constructed from an alignment of 1967 core genes constructed from the pangenome analysis of 311 Agrobacterium strains. The four main clades are highlighted in color. Purple: cluster 1 (C1; *A. tumefaciens*); Light blue: cluster 2 (C2; mixture of named and unclassified *Agrobacterium*); Green: cluster 3 (C3; *A. rhizogenes*); Pink: cluster 4 (C4; *A. vitis*). The outside ring demonstrates the 35 ANI groups.

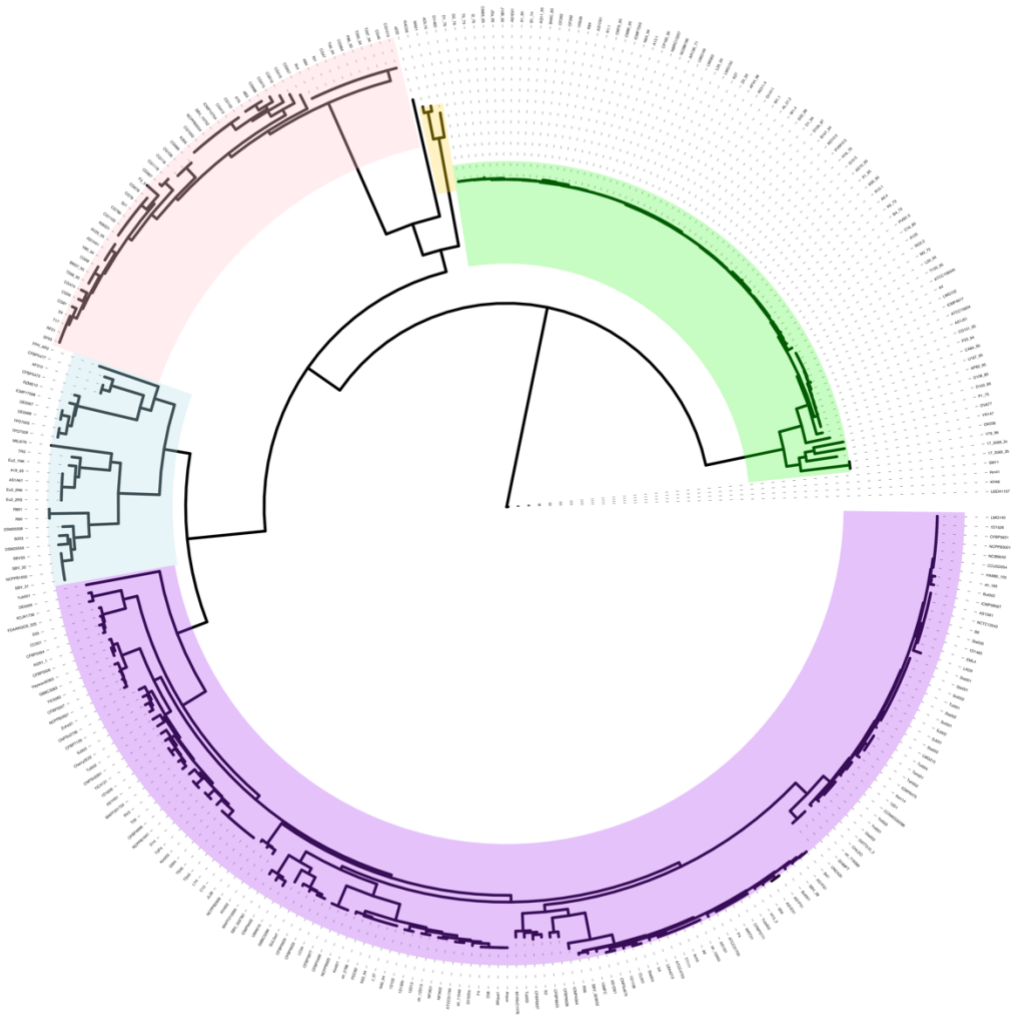


Fig. 2.4. Multi-locus phylogeny based on six house-keeping genes. The tree was rooted at the common ancestor of four *Sinorhizobium meliloti* strains. Four different colors highlighted four main clades. Purple: cluster 1 (C1; *A. tumefaciens*); Light blue: cluster 2 (C2; mixture of named and unclassified *Agrobacterium*); Green: cluster 3 (C3; *A. rhizogenes*); Pink: cluster 4 (C4; *A. vitis*).

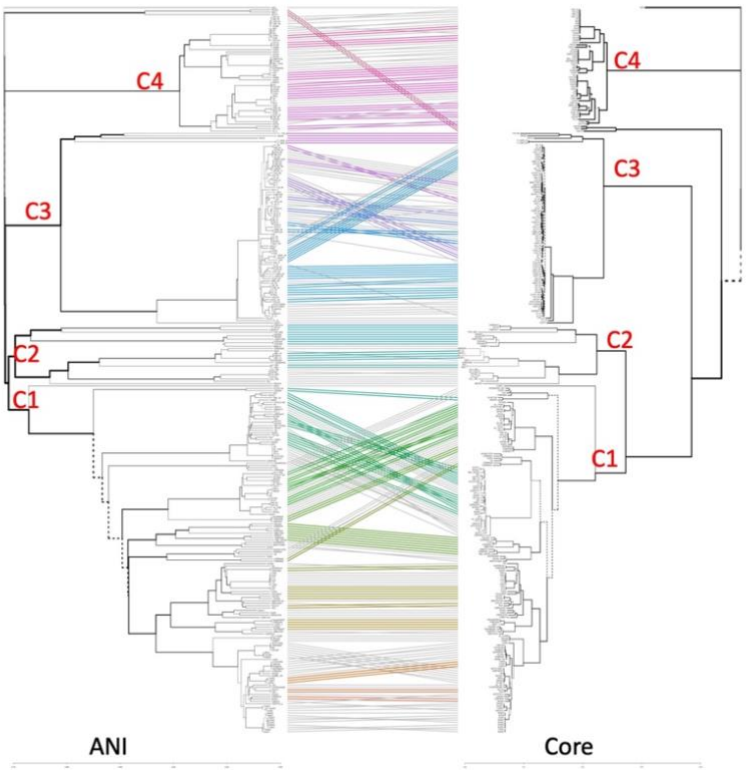


Fig. 2.5. Topology comparison of ANI-based (left) phylogeny and core gene (right) phylogeny of 311 *Agrobacterium* strains. Both trees were rooted at *A. tumefaciens* strain a222. Lines in the center with different colors highlight the distinct edges/branches of the two trees. The four clades in both trees are annotated as C1, C2, C3, and C4.

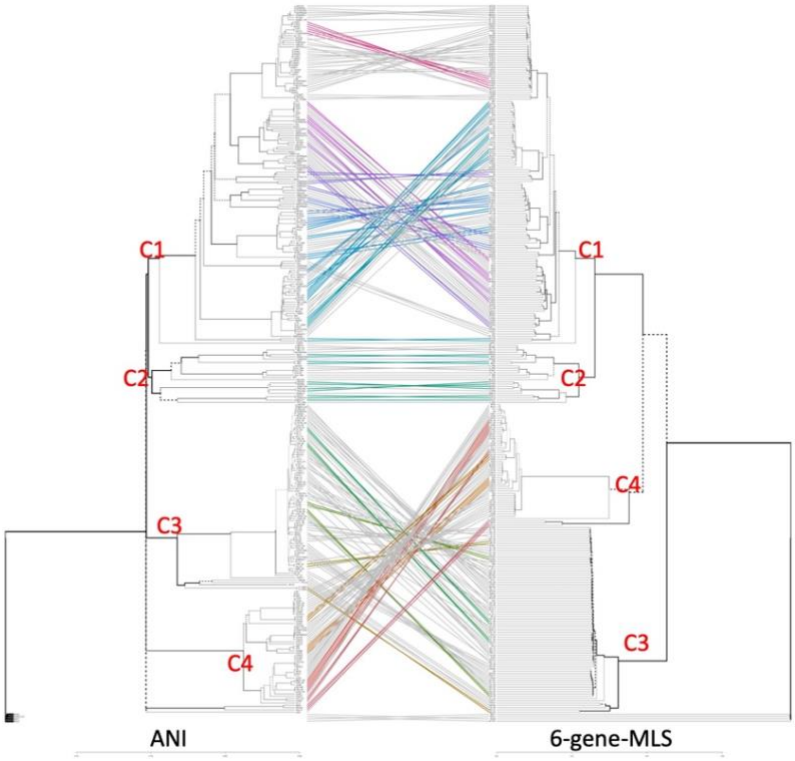


Fig. 2.6. Topology comparison of the ANI-based (left) phylogeny and six-gene phylogeny of 311 *Agrobacterium* strain. Both trees were rooted at the common ancestor of four *Sinorhizobium meliloti* strains. Lines in the center with different colors highlight the distinct edges/branches of the two trees. The four clades in both trees are annotated as C1, C2, C3, and C4.

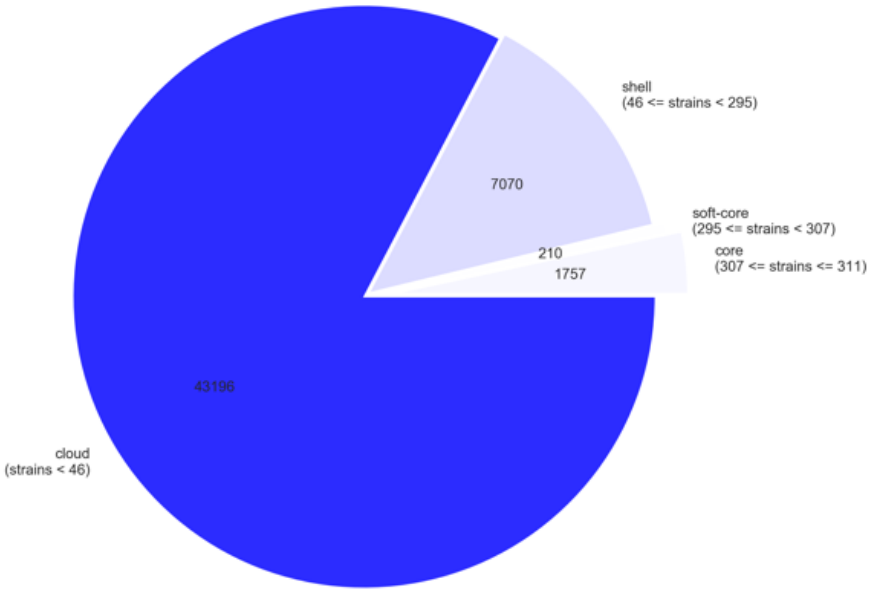


Fig. 2.7. Pangenome of 311 *Agrobacterium* strains. Core genes were present in more than 99% of the strains. Soft core genes were present in 95% to 99% of strains. Shell genes and cloud genes constitute the accessory genes. Shell genes were present in between 15% and 95% strains. Cloud genes were present in less than 15% of the strains examined.

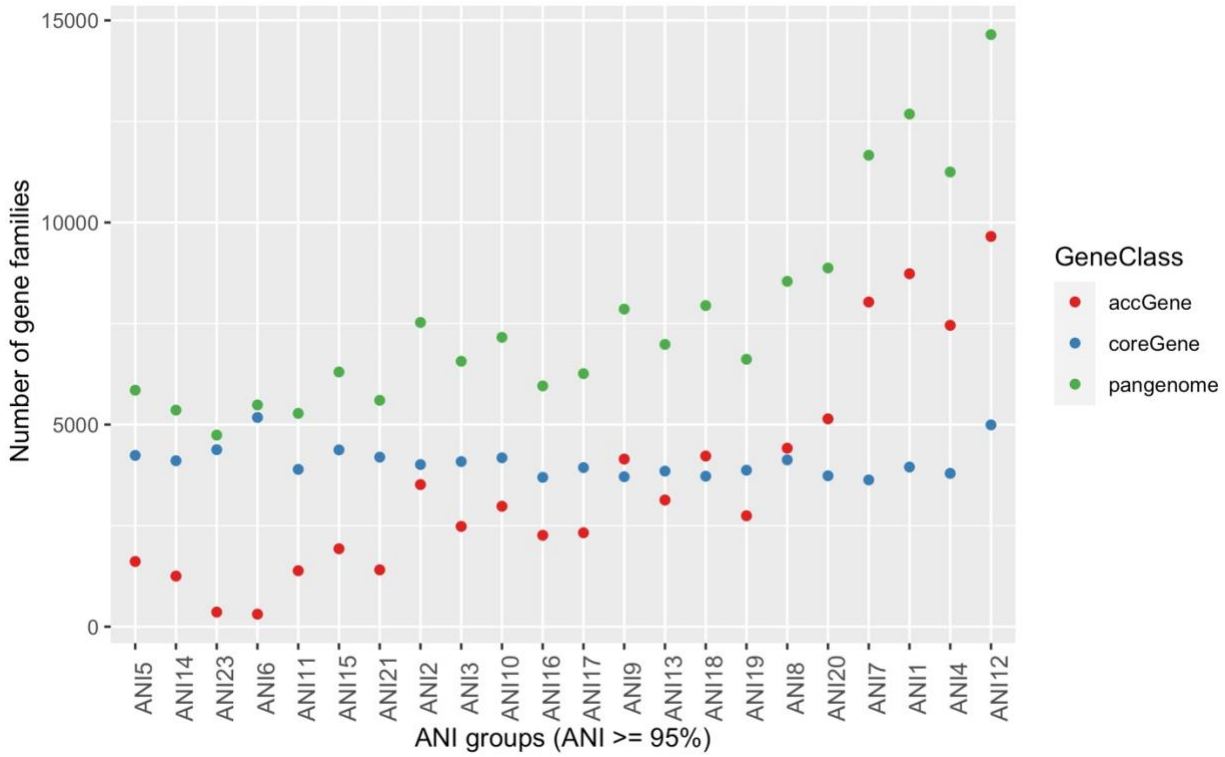


Fig. 2.8. Pangenome of each ANI group. The X-axis contains 22 ANI groups ordered by the number of strains in each ANI group from the least (ANI5 with 2 strains) to largest (ANI12 with 76 strains). The Y-axis represents the number of gene families. Green color: Pangenome size; Red color: Number of accessory gene families; Blue color: Number of core gene families.

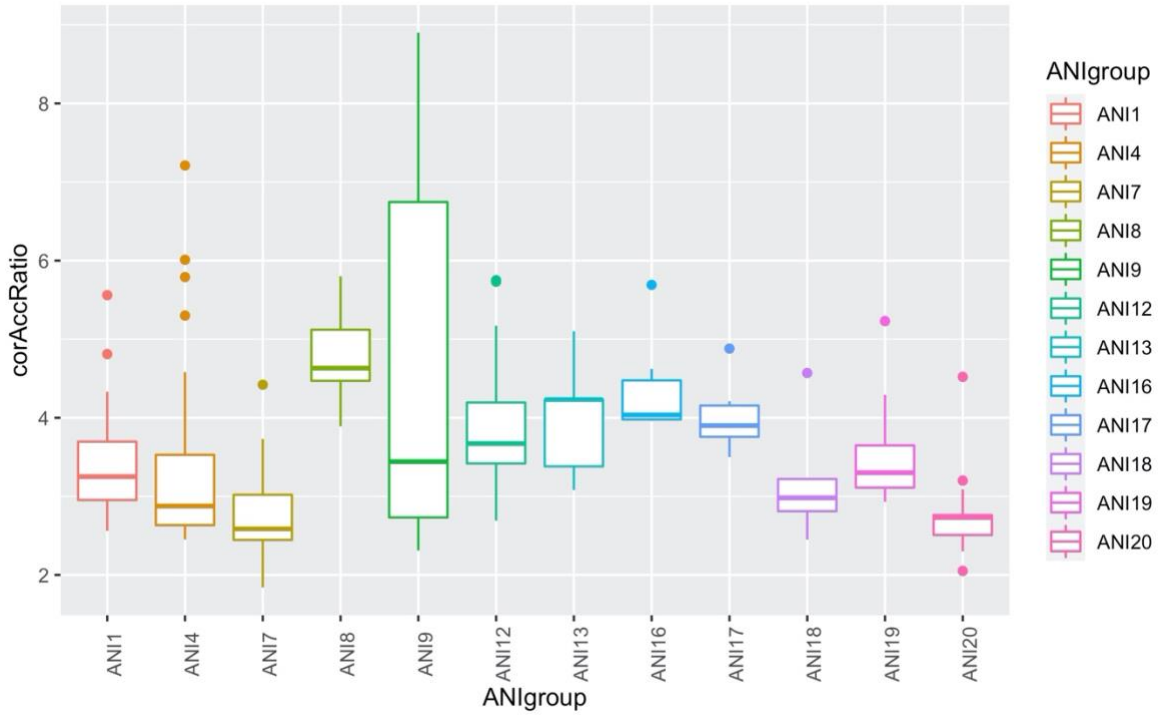


Fig. 2.9. Core-to-accessory gene ratio of ANI groups containing more than 5 strains.

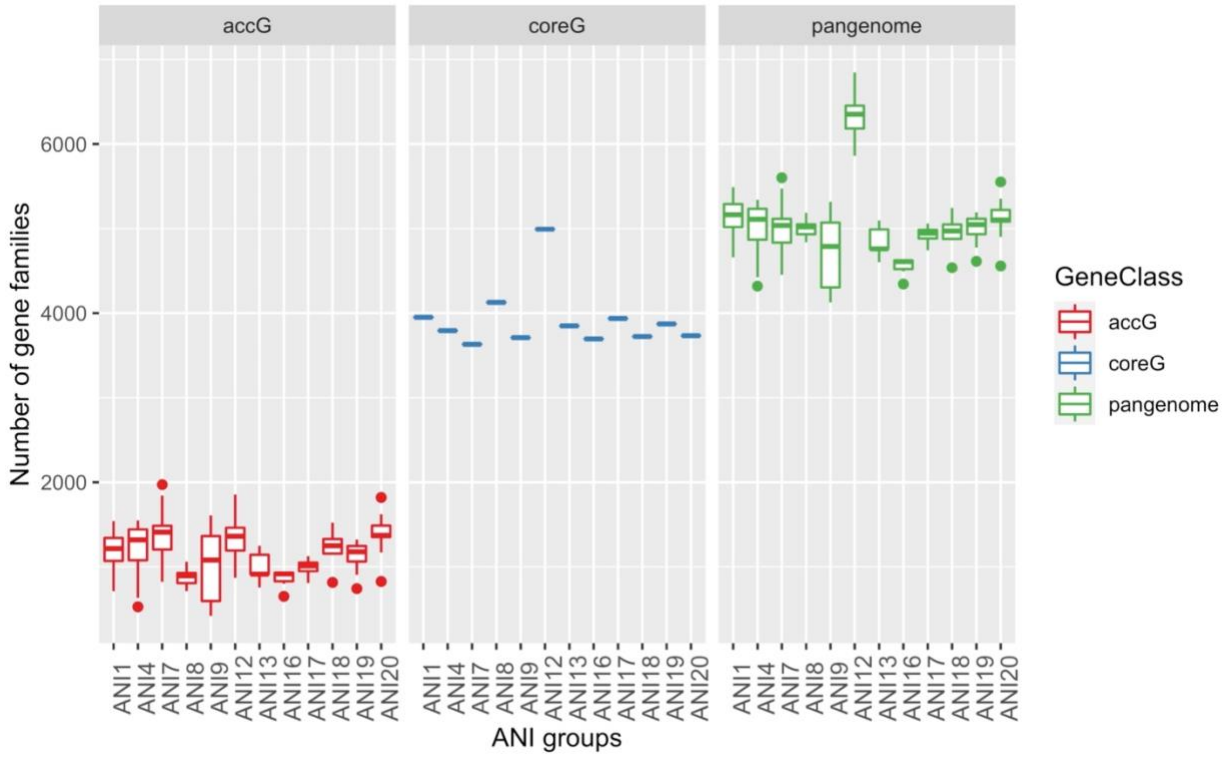


Fig. 2.10. Pangenome of each ANI group. X-axis indicates ANI group name. Y-axis indicates number of gene families. Red color: accessory genes; Blue color: core genes; Green color: pangenome.

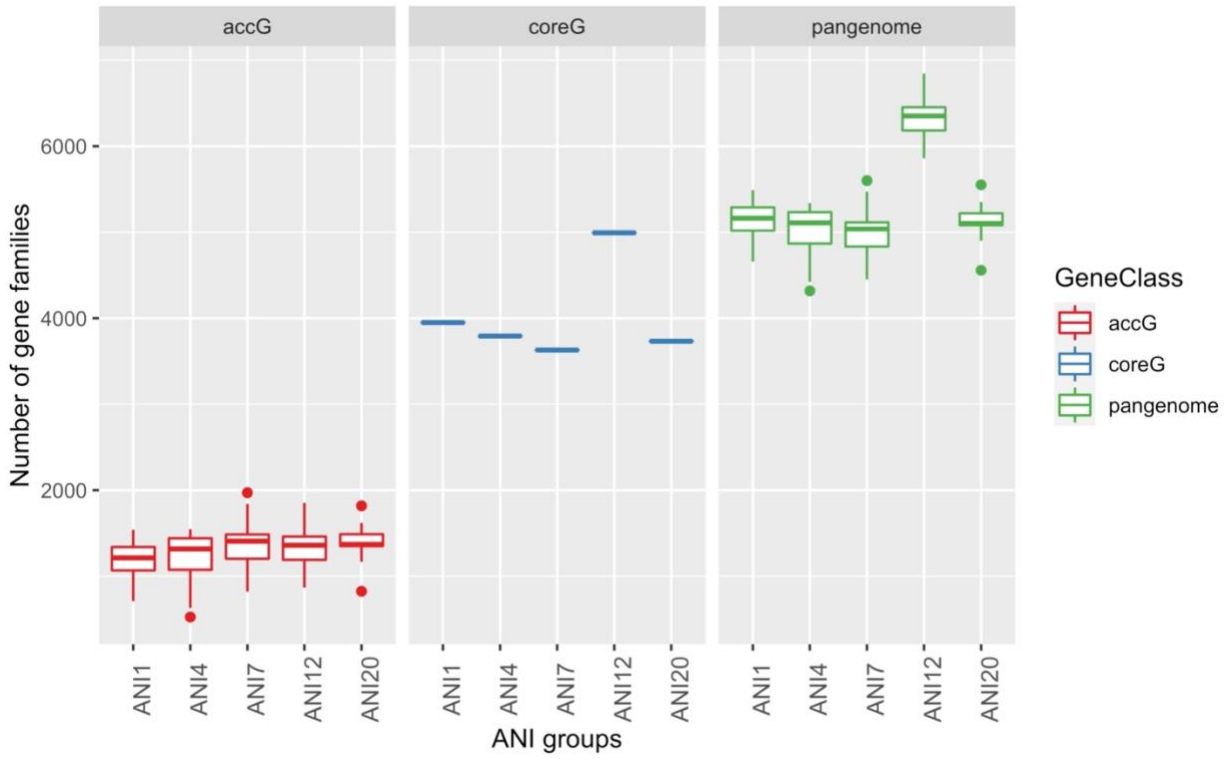


Fig. 2.11. Pangenome of ANI groups containing more than 20 strains. X-axis indicates ANI group name. Y-axis indicates number of gene families. Red color: accessory genes; Blue color: core genes; Green color: pangenome.

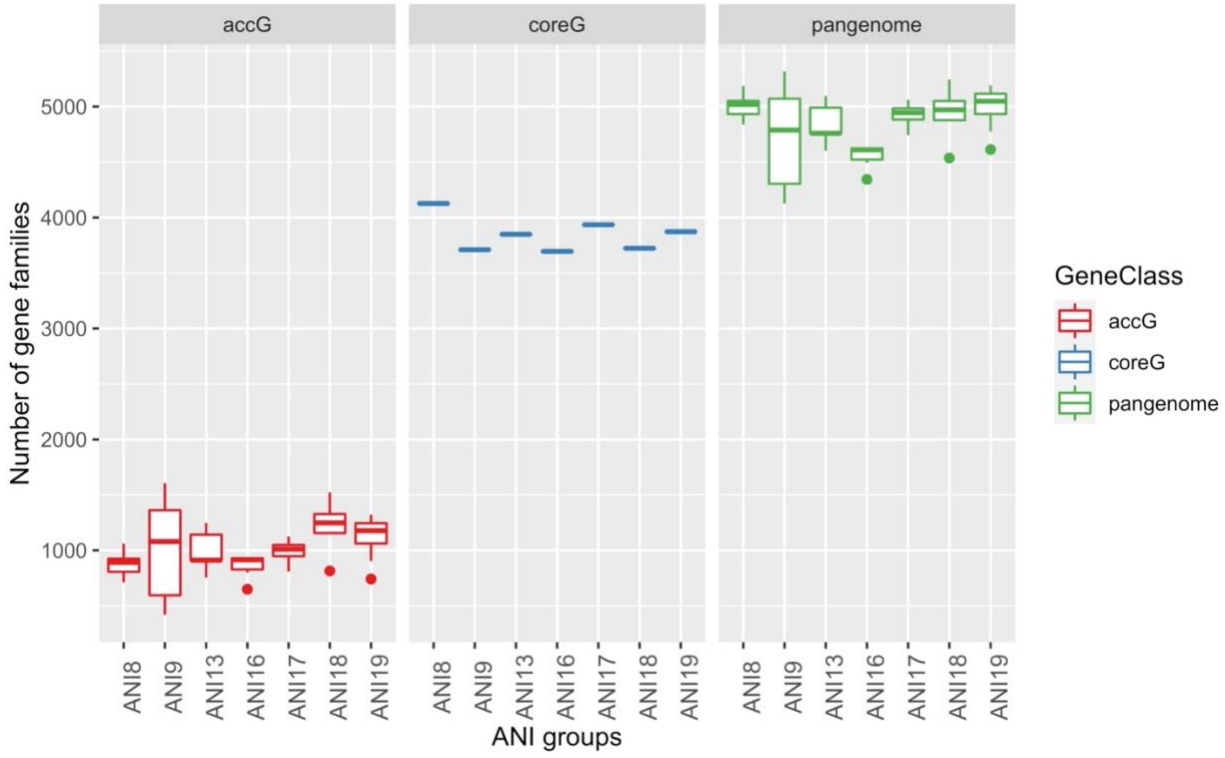


Fig. 2.12. Pangenome of ANI groups containing 6-19 strains. X-axis indicates ANI group name. Y-axis indicates number of gene families. Red color: accessory genes; Blue color: core genes; Green color: pangenome.

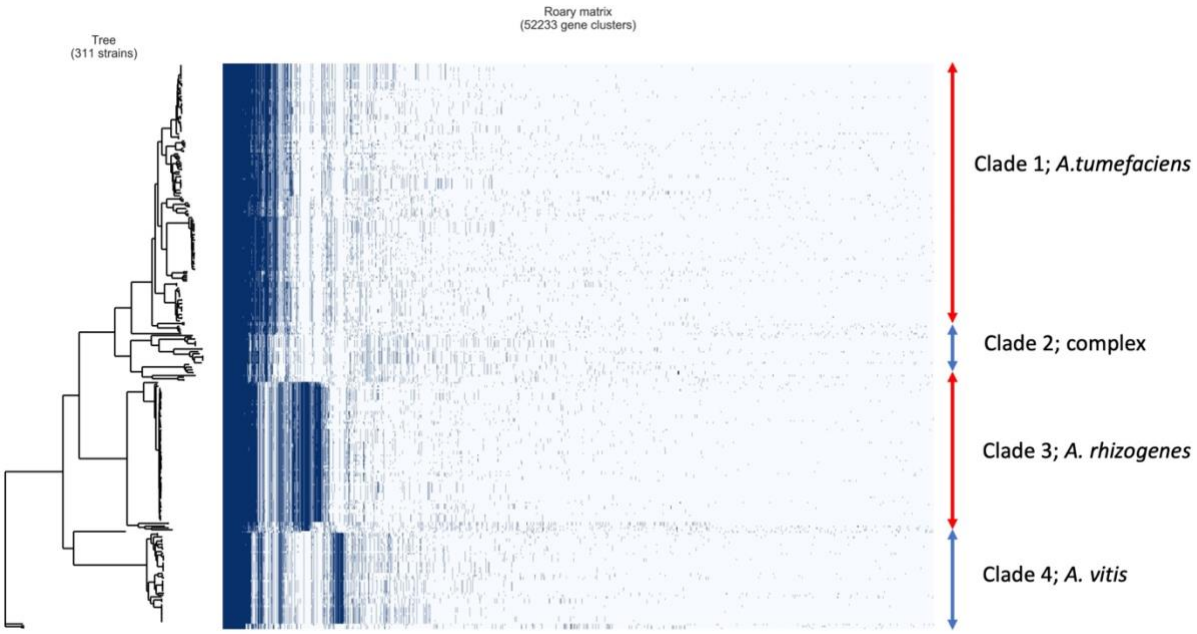


Fig. 2.13. Phylogenetic profile with a presence/absence matrix of all genes in the pangenome. Color blue indicates the presence of a gene in a strain. Color white indicates the absence of a gene in a strain. Phylogenetic profile with a presence/absence matrix of all genes in the pangenome. The four clades are annotated on the right. The gene presence/absence profile supports the 4 clades recovered from Core phylogeny.

Table 2.1. Thirty-five ANI groups with their corresponding NCBI taxonomy. ANI groups were named based on NCBI taxonomy. *A. tumefaciens* strains the genomic species group designation can be found within the parenthesis.

ANI Groups	Number of Strains	NCBI Taxonomy
ANI1	30	*A. tumefaciens* (G1)
ANI2	4	*A. tumefaciens* (G2)
ANI3	4	*A. tumefaciens* (G3)
ANI4	42	*A. tumefaciens* (G4)
ANI5	2	*A. tumefaciens* (G5)
ANI6	3	*A. tumefaciens* (G6)
ANI7	24	*A. tumefaciens* (G7)
ANI8	19	*A. tumefaciens* (G8)
ANI9	7	*A. tumefaciens* (G9)
ANI10	4	*A. tumefaciens* (G13)
ANI11	3	*A. larrymoorei*
ANI12	76	*A. rhizogenes*
ANI13	7	*A. rosae*
ANI14	2	*A. tumefaciens*
ANI15	3	*A. tumefaciens*
ANI16	6	*A. fabrum/A.tumefaciens/A.rhizogenes/Agrobacterium spp.*
ANI17	6	*A. rubi*
ANI18	13	*A. vitis*
ANI19	14	*A. vitis*
ANI20	22	*A. vitis*
ANI21	3	*A. tumefaciens*
ANI22	2	*A. bohemicum*
ANI23	2	*Agrobacterium spp.*
ANI24	2	*A. rhizogenes*
ANI25	1	*A. rhizogenes*
ANI26	1	*Agrobacterium spp.*
ANI27	1	*A. albertmagni*
ANI28	1	*A. vitis*
ANI29	1	*A. tumefaciens*
ANI30	1	*A. larrymoorei*
ANI31	1	*A. rhizogenes*
ANI32	1	*A. rhizogenes*
ANI33	1	*A. rhizogenes*
ANI34	1	*A. rhizogenes*
ANI35	1	*A. rhizogenes*

Table 2.2. Pangenome summary of 311 Agrobacterium strains. The % isolates column is the percentage of isolates with the characteristics indicated for that row. The # of clusters indicates the number of gene families. The >1 allele column shows number of gene families containing more than one allele. Last two columns: number of gene families containing a gene fission/fusion; number of multicopy gene families.

%isolates	#clusters	>1 allele	fission/fusion	multicopy
0-10%	42290	1859	650	167
10-25%	3061	790	294	144
25-50%	3207	610	408	163
50-75%	1101	293	286	124
75-90%	503	162	158	100
90-95%	104	25	39	23
95-100%	1967	246	506	165

Table 2.3. Pangenome of ANI groups. This table shows the pangenome, core genome, and accessory genome sizes as the number of gene families in each ANI group.

ANI group	Number of Strains	Pangenome Size	Core Genome	Accessory Genome	Core/acc ratio	Core/pangenome ratio
ANI1	30	12682	3950	8732	0.45	0.31
ANI2	4	7527	4011	3516	1.14	0.53
ANI3	4	6566	4085	2481	1.65	0.62
ANI4	42	11249	3793	7456	0.51	0.34
ANI5	2	5851	4238	1613	2.63	0.72
ANI6	3	5486	5176	310	16.7	0.94
ANI7	24	11663	3631	8032	0.45	0.31
ANI8	19	8544	4127	4417	0.93	0.48
ANI9	7	7857	3710	4147	0.89	0.47
ANI10	4	7158	4179	2979	1.4	0.58
ANI11	3	5277	3892	1385	2.81	0.74
ANI12	76	14648	4993	9655	0.52	0.34
ANI13	7	6983	3849	3134	1.23	0.55
ANI14	2	5358	4106	1252	3.28	0.77
ANI15	3	6302	4374	1928	2.27	0.69
ANI16	6	5957	3695	2262	1.63	0.62
ANI17	6	6260	3936	2324	1.69	0.63
ANI18	13	7945	3723	4222	0.88	0.47
ANI19	14	6616	3872	2744	1.41	0.59
ANI20	22	8874	3733	5141	0.73	0.42
ANI21	3	5601	4195	1406	2.98	0.75
ANI22	2	4328	4280	48	89.17	0.99
ANI23	2	4740	4379	361	12.13	0.92
ANI24	2	7211	7204	7	1029.14	1

References

1. Conn, H. J. Validity of the Genus Alcaligenes. *J. Bacteriol.* **44**, 353–360 (1942).
2. Tzfira, T. & Citovsky, V. *Agrobacterium: From Biology to Biotechnology*. (Springer Science & Business Media, 2007).
3. Kado, C. I. Historical account on gaining insights on the mechanism of crown gall tumorigenesis induced by *Agrobacterium tumefaciens*. *Front. Microbiol.* **5**, (2014).
4. de Cleene, M. Crown gall: economic importance and control. *Zentralblatt Bakteriol. Parasitenkd. Infekt. Hyg. Zweite Naturwissenschaftliche Abt. Mikrobiol. Landwirtsch. Technol. Umweltschutzes* **134**, 551–554 (1979).
5. Mansfield, J. *et al.* Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* **13**, 614–629 (2012).
6. Gelvin, S. B. *Agrobacterium Biology: From Basic Science to Biotechnology*. (Springer, 2018).
7. Young, J. M., Kuykendall, L. D., Martínez-Romero, E., Kerr, A. & Sawada, H. Classification and nomenclature of *Agrobacterium* and *Rhizobium*. *Int. J. Syst. Evol. Microbiol.* **53**, 1689–1695 (2003).
8. Holmes, B. & Roberts, P. The Classification, Identification and Nomenclature of *Agrobacteria*. *J. Appl. Bacteriol.* **50**, 443–467 (1981).
9. Ophel, K. & Kerr, A. Agrobacterium vitis sp. nov. for Strains of *Agrobacterium* biovar 3 from Grapevines. *Int. J. Syst. Evol. Microbiol.* **40**, 236–241 (1990).
10. Genetello, C. *et al.* Ti plasmids of *Agrobacterium* as conjugative plasmids. *Nature* **265**, 561–563 (1977).
11. Mousavi, S. A. *et al.* Phylogeny of the *Rhizobium–Allorhizobium–Agrobacterium* clade supports the delineation of Neorhizobium gen. nov. *Syst. Appl. Microbiol.* **37**, 208–215 (2014).
12. Young, J. M., Kuykendall, L. D., Martínez-Romero, E., Kerr, A. & Sawada, H. A revision of Rhizobium Frank 1889, with an emended description of the genus, and the inclusion of all species of

Agrobacterium Conn 1942 and *Allorhizobium undicola* de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes, R. rubi, R. undicola* and *R. vitis*. *Int. J. Syst. Evol. Microbiol.* **51**, 89–103 (2001).

13. Kisand, V. & Wikner, J. Limited resolution of 16S rDNA DGGE caused by melting properties and closely related DNA sequences. *J. Microbiol. Methods* **54**, 183–191 (2003).
14. STACKEBRANDT, E. & GOEBEL, B. M. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *Int. J. Syst. Evol. Microbiol.* **44**, 846–849 (1994).
15. Janda, J. M. & Abbott, S. L. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. *J. Clin. Microbiol.* **45**, 2761–2764 (2007).
16. Farrand, S. K., van Berkum, P. B. & Oger, P. *Agrobacterium* is a definable genus of the family Rhizobiaceae. *Int. J. Syst. Evol. Microbiol.* **53**, 1681–1687 (2003).
17. Jumas-Bilak, E., Michaux-Charachon, S., Bourg, G., Ramuz, M. & Allardet-Servent, A. Unconventional Genomic Organization in the Alpha Subgroup of the Proteobacteria. *J. Bacteriol.* **180**, 2749–2755 (1998).
18. Costechareyre, D. *et al.* Rapid and efficient identification of *Agrobacterium* species by recA allele analysis: *Agrobacterium* recA diversity. *Microb. Ecol.* **60**, 862–872 (2010).
19. Aujoulat, F. *et al.* Multilocus Sequence-Based Analysis Delineates a Clonal Population of *Agrobacterium (Rhizobium) radiobacter (Agrobacterium tumefaciens)* of Human Origin ▿. *J. Bacteriol.* **193**, 2608–2618 (2011).
20. Revised phylogeny of Rhizobiaceae: Proposal of the delineation of Pararhizobium gen. nov., and 13 new species combinations. *Syst. Appl. Microbiol.* **38**, 84–90 (2015).
21. Weisberg, A. J. *et al.* Unexpected conservation and global transmission of agrobacterial virulence plasmids. *Science* **368**, (2020).

22. History and current taxonomic status of genus *Agrobacterium* | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S0723202019303418?token=67EABBBB09BF51A46DA419D4F4C3C7544F55003A4545CA8C9909CB7DF02495F6B051A7CBCBA7CD8DF5FDE1092F1C806B doi:10.1016/j.syapm.2019.126046.
23. Tindall, B. J., Rosselló-Móra, R., Busse, H.-J., Ludwig, W. & Kämpfer, P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* **60**, 249–266 (2010).
24. Konstantinidis, K. T. & Tiedje, J. M. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2567–2572 (2005).
25. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
26. Orata, F. D., Meier-Kolthoff, J. P., Sauvageau, D. & Stein, L. Y. Phylogenomic Analysis of the Gammaproteobacterial Methanotrophs (Order Methylococcales) Calls for the Reclassification of Members at the Genus and Species Levels. *Front. Microbiol.* **9**, (2018).
27. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
28. Romalde, J. L., Balboa, S. & Ventosa, A. Editorial: Microbial Taxonomy, Phylogeny and Biodiversity. *Front. Microbiol.* **10**, (2019).
29. Konstantinidis, K. T., Ramette, A. & Tiedje, J. M. The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **361**, 1929–1940 (2006).
30. Costechareyre, D., Bertolla, F. & Nesme, X. Homologous Recombination in *Agrobacterium*: Potential Implications for the Genomic Species Concept in Bacteria. *Mol. Biol. Evol.* **26**, 167–176 (2009).
31. Portier, P. *et al.* Identification of Genomic Species in *Agrobacterium* Biovar 1 by AFLP Genomic Markers. *Appl. Environ. Microbiol.* **72**, 7123–7131 (2006).

32. Riley, M. A. & Lizotte-Waniewski, M. Population Genomics and the Bacterial Species Concept. in *Horizontal Gene Transfer: Genomes in Flux* (eds. Gogarten, M. B., Gogarten, J. P. & Olendzenski, L. C.) 367–377 (Humana Press, 2009). doi:10.1007/978-1-60327-853-9_21.

33. Tettelin, H. *et al.* Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci.* **102**, 13950–13955 (2005).

34. Rouli, L., Merhej, V., Fournier, P.-E. & Raoult, D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* **7**, 72–85 (2015).

35. Ku, C. *et al.* Endosymbiotic gene transfer from prokaryotic pangenomes: Inherited chimerism in eukaryotes. *Proc. Natl. Acad. Sci.* **112**, 10139–10146 (2015).

36. Tomida, S. *et al.* Pan-Genome and Comparative Genome Analyses of Propionibacterium acnes Reveal Its Genomic Diversity in the Healthy and Diseased Human Skin Microbiome. *mBio* **4**, (2013).

37. Vinuesa, P. & Contreras-Moreira, B. Pangenomic Analysis of the Rhizobiales Using the GET_HOMOLOGUES Software Package. in (2015). doi:10.1002/9781119053095.CH27.

38. González, V. *et al.* Phylogenomic *Rhizobium* Species Are Structured by a Continuum of Diversity and Genomic Clusters. *Front. Microbiol.* **10**, (2019).

39. Sedlazeck, F. J., Lee, H., Darby, C. A. & Schatz, M. C. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat. Rev. Genet.* **19**, 329–346 (2018).

40. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

41. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput. Biol.* **13**, e1005595 (2017).

42. Zhao, Y. *et al.* PGAP: pan-genomes analysis pipeline. *Bioinformatics* **28**, 416–418 (2012).

43. Zhao, Y. *et al.* PGAP-X: extension on pan-genome analysis pipeline. *BMC Genomics* **19**, 36 (2018).

44. Poret-Peterson, A. T., Bhatnagar, S., McClean, A. E. & Kluepfel, D. A. Draft Genome Sequence of *Agrobacterium tumefaciens* Biovar 1 Strain 186, Isolated from Walnut. *Genome Announc.* **5**, (2017).
45. Gascuel, O. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* **14**, 685–695 (1997).
46. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
47. Bayliss, S. C., Thorpe, H. A., Coyle, N. M., Sheppard, S. K. & Feil, E. J. PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. *GigaScience* **8**, (2019).
48. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* **5**, e9490 (2010).
49. Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
50. Caputo, A. *et al.* Pan-genomic analysis to redefine species and subspecies based on quantum discontinuous variation: the Klebsiella paradigm. *Biol. Direct* **10**, 55 (2015).
51. Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
52. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
53. Guillerme, T. dispRity: A modular R package for measuring disparity. *Methods Ecol. Evol.* **9**, 1755–1763 (2018).
54. dispRity.pdf.
55. Guillerme, T. & Cooper, N. Time for a rethink: time sub-sampling methods in disparity-through-time analyses. *Palaeontology* **61**, 481–493 (2018).
56. Welcome | ggplot2. https://ggplot2-book.org/.

57. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

58. Wang, L.-G. *et al.* Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Mol. Biol. Evol.* **37**, 599–603 (2020).

59. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

60. Wickham, H. & RStudio. *tidyr: Tidy Messy Data*. (2020).

61. Henry, L., Wickham, H. & RStudio. *tidyselect: Select from a Set of Strings*. (2020).

62. Ng, R. K. and P. *SparseM: Sparse Linear Algebra*. (2019).

63. Wickham, H., Seidel, D. & RStudio. *scales: Scale Functions for Visualization*. (2020).

64. Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).

65. Ridé, M. *et al.* Characterization of Plasmid-Borne and Chromosome-Encoded Traits of Agrobacterium Biovar 1, 2, and 3 Strains from France. *Appl. Environ. Microbiol.* **66**, 1818–1825 (2000).

66. Kuzmanović, N. *et al.* Revisiting the taxonomy of *Allorhizobium vitis* (i.e., *Agrobacterium vitis*) using genomics - emended description of *All. vitis* sensu stricto and description of *Allorhizobium ampelinum* sp. nov. *bioRxiv* 2020.12.19.423612 (2020) doi:10.1101/2020.12.19.423612.

67. Lassalle, F. *et al.* Genomic species are ecological species as revealed by comparative genomics in *Agrobacterium tumefaciens*. *Genome Biol. Evol.* **3**, 762–781 (2011).

68. Caputo, A., Fournier, P.-E. & Raoult, D. Genome and pan-genome analysis to classify emerging bacteria. *Biol. Direct* **14**, 5 (2019).

69. Rouli, L., Merhej, V., Fournier, P.-E. & Raoult, D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* **7**, 72–85 (2015).

Chapter 3

Phenotypic diversity of *Agrobacterium tumefaciens* strains isolated from walnut orchards across the central valley of California

Abstract

Crown gall is one of the major soil-borne diseases posing a significant threat to the walnut industry in California orchards. The causal agent of this disease is *Agrobacterium tumefaciens*, which has an extremely wide host range that includes most dicotyledonous plants [1]. The mechanism of pathogenesis of this bacterium has been extensively investigated [2], however, managing the increasing incidence of crown gall disease on walnut is still a challenge. Here, I examine 28 *A. tumefaciens* strains isolated from crown gall and soil samples collected from the top 10 walnut growing counties in California, as well as two additional *A. tumefaciens* strains, C58 and Chilean strain CL001. Phenotypic traits of these strains were determined including virulence on *Datura stramonium* (*D. stramonium*) and walnut plants, antibiotic sensitivity, resistance to the biocontrol strain K84, growth rates in rich and poor medium, and motility *in vitro*. The virulence testing on *D. stramonium* and walnut revealed considerable differences in virulence among the 30 strains. The doubling time ranged from 1 h to 4 h in the nutrient rich medium, whereas the doubling time in the nutrient poor medium ranged from 1.5 h to 10.5 h. Antibiotic resistance profiles showed that most strains were sensitive to chloramphenicol, carbenicillin, erythromycin, ciprofloxacin, rifampin, and tetracycline. Interestingly, most strains were resistant to streptomycin and vancomycin, which are broadly used in agriculture. Nearly half of the strains were resistant to K84, which is used as a biocontrol for crown gall of walnut. The motility test revealed significant differences among the 30 strains. These data provide insights into the phenotypic diversity of *A. tumefaciens*, which may be exploited in the development of sustainable crown gall management strategies.

Keywords: Agrobacterium tumefaciens, virulence, motility, antibiotic resistance, growth rate, K84

Introduction

Crown gall development is the result of a complex interaction between *A. tumefaciens* and a susceptible plant host. A plant wound, which is a prerequisite for tumor formation, releases acetosyringone (AS), which is sensed by the VirA/VirG two components regulatory system of *A.*

tumefaciens [3]. In the presence of elevated AS concentrations, this two component regulatory system initiates expression of the *vir* operon [3]. After initial bacterial attachment, T-DNA is transferred to plant cells through a type IV secretion system, where it is stably inserted into the plant genome [3]. Expression of the plant derived hormone genes contained in the T-DNA results in uncontrolled hormone production, which leads to undifferentiated plant cell proliferation, i.e., tumor formation. At the same time, opine biosynthesis genes contained on the same T-DNA produce opines, which serve as carbon and nitrogen sources for *A. tumefaciens* strains harboring opine-specific catabolism genes on the Ti plasmid [3]. Opines also promote conjugation and transfer of the Ti plasmid among *A. tumefaciens* cells, resulting in a greater percentage of the *A. tumefaciens* population that can use opines and transform plant cells [2]. Modification or changes to any of these steps may have significant impacts on pathogenesis. Whereas the process of pathogenesis of *Agrobacterium* infection is fully understood, there exists limited information on the phenotypic diversity of *Agrobacterium* strains collected across a geographic gradient in an economically important crop.

The first step in tumorigenesis requires the pathogen to reach and colonize the host surface. Various motility types have been identified as essential in the process of host surface colonization [4]. These includes swimming, swarming, gliding, twitching, sliding, and darting [4]. Motility also facilitates the ability of microorganisms to actively move in soil or other environments. Generally, quantitative mobility comparisons performed under laboratory conditions have been shown to be valid predictors of bacterial movement in the environment [5]. For example, the swimming motility of *Ralstonia solanacearum* measured *in vitro* was required for tomato plant invasion and colonization [6]. The swimming motility of *A. tumefaciens* strains is also important for biofilm formation and host surface colonization, which are driven by flagella [7]. Motility and chemotaxis are also essential in the bacterial life cycle and are intimately linked with virulence via complex regulatory networks [8]. Studying motility of pathogenic tumor-inducing *A. tumefaciens* strains will provide insights into the role of this trait in ecology and virulence. Therefore,

motility of pathogens may serve as a factor to help predict the disease epidemiology in the future. It also may serve as a specific target to facilitate disease management [8].

Pathogen population density is an important factor in causing plant disease [9]. Monitoring the plant pathogen population density often facilitates epidemiological predictions [9]. The pathogen growth rate also serves as an important ecological trait of a bacterial species, and is a useful indicator of ecological fitness [10]. The typical growth curve of a bacterial population includes a lag phase, exponential phase, stationary phase, and death phase [11]. The growth rate (i.e., doubling time) of a bacterial population is normally calculated during the exponential phase of the growth curve [10]. Because the population density of a pathogen serves as an important factor in disease incidence [12], comparing doubling time is a useful approach to assessing ecological fitness, whereas providing epidemiological insights into this group of *A. tumefaciens* strains isolated from different geographic region.

Not only is bacterial population density attainment related to growth rate, it also contributes to bacterial sensitivity to antibiotics [13]. For example, as population density increases, in general, biofilm formation is enhanced, which frequently results in elevated levels of antibiotic resistance [14]. Antibiotics or antimicrobial drugs that are cytotoxic or cytostatic to microorganisms, evolved in nature long before being discovered by scientists [15]. They serve as “weapons” produced by members of the community to facilitate competition for ecological niches or resources. All members of microbial communities fight for limited resources to survive in these highly competitive surroundings [15].

Antibiotics have been broadly used in managing human, animal, and plant diseases [16]. However, the broad use of antibiotics has led to development of antibiotic resistance, which occurs when antibiotic resistant members of the population are selected. The origins of antibiotic-resistance in members of a given bacterial population vary, and may include acquisition of existing resistance genes from the environmental microbiome or perhaps selection of spontaneous mutants that are antibiotic-resistant [17]. The first approach can be described as horizontal gene transfer (HGT), which is a global concern in human

health, animal husbandry, agriculture, and the environment [17]. Antibiotic resistance has positively correlated with the level and improper use of antibiotics [17]. Given the HGT of antibiotic resistance genes, it is important to evaluate the development of antibiotic resistance in plant pathogens including *A. tumefaciens*. In vitro, antibiotic resistance tests provide resistance profiles, and insights into understanding how *A. tumefaciens* strains may respond to other microbes secreting corresponding antibiotics in the soil. For the 28 *Agrobacterium* strains from California walnuts examined, I assessed resistance to eight antibiotics exhibiting different modes of action in order to evaluate the widespread nature of antibiotic resistance in *A. tumefaciens*.

Biocontrol strains of bacteria that secret antibiotics have been selected from nature and developed into commercial products for disease management [18]. NOGALL and GALLTROL-A are the main commercial biocontrol products for controlling *A. tumefaciens*, including in walnuts [18]. The active agent in NOGALL is *Agrobacterium* strain K84, whereas GALLTROL-A contains *A. tumefaciens* strain K1026. The latter is a *tra*-deletion mutant of pAgK84 [19], which prevents pAgK84 transfer from strain K84 to pathogenic *A. tumefaciens* strains that has led to resistant to K84. Furthermore, only strains with agrocinopine-type Ti plasmid are sensitive to K84 [20], which restricts the utility of this biocontrol method to agrocinopine *Agrobacterium* strains. However, knowledge concerning K84 sensitivity of pathogenic *A. tumefaciens* strains isolated from California walnut orchards is lacking. The data presented here provide information on the range of K84 sensitivity of pathogenic *A. tumefaciens* strains isolated from commercial walnut orchards or nurseries. This K84 efficacy data will facilitate grower decisions on the utility of K84 under commercial conditions.

In the present study, I selected virulence level, growth rate, motility, antibiotic resistance, and K84 sensitivity as five key phenotypic traits to assess diversity of *Agrobacterium tumefaciens* strains isolated from walnut orchards in California. I also examined the potential genetic bases underlying these five traits in Chapter 4.

Methods

Isolation and growth of *Agrobacterium tumefaciens* strains

Twenty-eight *A. tumefaciens* strains were isolated from fresh tumors (green, soft tumors), wood stem disk, or soil samples collected from walnut orchards or nurseries located in the top ten walnut growing counties in California from 2004-2019. Strain CL001 was isolated from a soil sample of a walnut orchard collected in Chile in 2017. Tumors or wood disks were first washed with running tap water, rinsed in deionized (DI) water and dried with paper towels in a laminar flow hood. Samples were then surface disinfested by spraying 70% ethanol for 30 seconds and then allowed to dry in the laminar flow hood. A sterile blade was used to remove and discard the outside layer of the tumor. Small pieces of the tumor interior were placed on the surface of semi-selective 1A medium and incubated for three days at 28 °C.

The candidate *A. tumefaciens* colony is dome-shaped and purple when growing on 1A medium. Each putative *A. tumefaciens* colony was streaked on tryptic soy broth agar (TSBA) with each plate having 10 short streaks followed by incubating overnight at 28 °C. Each of the short streaks was transferred into a 24-well plate containing lactose agar medium with each strain in one well and incubated at 28 °C for 24 h. The positive control was *A. tumefaciens*, 186 strain, and *Brenneria nigrifluens* strain as a negative control. These putative *A. tumefaciens* strains were then subjected to Benedict's test. Because the A. tumefaciens can convert lactose to 3-ketolactose, making the blue reagent turn to yellow color. First, 100 microliter Benedict's reagents were added to each well and the plate was incubated at 28 °C for one h. The wells that turned yellow indicated that the original colonies were *A. tumefaciens* strains. The well containing the negative control remained blue color. The well containing the positive control turned yellow. The well position indicating positive for *A. tumefaciens* were then streaked onto fresh TSBA media and incubated for two d at 28 °C. Single colonies were selected for the tip6 colony PCR test to detect Ti-plasmid [21]. The primer pair Tip6F/R sequences are Tip6F 5'-GGTCTAATGCGCAGAGGTGT-3' and Tip6R 5'-CGGCTCAAGGATTAGACAGG-3'. The tip6 PCR direct the amplification of a 243 bp amplicon of the

intergenic region between gene 5 and *tms2*/*iaaH* of the T-DNA [21]. Finally, Tip6 PCR-positive strains were tested for virulent *A. tumefaciens*. Strain K84 and C58 were from my lab collection stored at -80 °C.

Isolation of *A. tumefaciens* from soil

One gram of soil was placed in a 50 mL conical screen cap sterile tube, to which 10 mL sterile water was added. The tube was then shaken at 200 rpm for 10 mins on a gyratory shaker. The suspension was then diluted in sterile water in a 10-fold serial dilution (10^{-1}, 10^{-2}, 10^{-3}, 10^{-5}, 10^{-6}, 10^{-7},10^{-8}, 10^{-9}), and spread onto 1A plates using an EDDY JET2 (Neutec Group INC) spiral plater in E-mode. The plates were incubated at 28 °C for three days. Putative *A. tumefaciens* colonies as described above were selected and transferred to fresh 1A plates. These cultures were grown for 2 days followed by storage at -80 °C as described below.

Culture preservation and preparation

Single colonies of 30 strains were grown in two mL liquid TSB overnight. Fifty microliters of this overnight culture of each strain were added to 50 microliters of a 30% glycerol/water solution, mixed, and stored at -80 °C. Ten tubes were prepared for each strain. For all the phenotypic tests, I first picked one -80°C storage tube and streaked the suspension onto TSBA and grown for 48 h to obtain single colonies. Then, a single colony of each strain was transferred to two mL TSB liquid and grown overnight. The overnight culture was then adjusted to a cell density of 10^6 cells mL^{-1} for phenotypic traits testing.

Plant materials and growth condition

Datura stramonium seedlings, grown from seed gifted by Andrew Hutchison in the UC Davis Plant Pathology Department, were maintained in Environmental, Horticulture, and Science (EHS) greenhouse at UC Davis under the following conditions: 16 h day length with supplemental lighting from 5-9 AM and 5-9 PM. Plants were watered twice a day (167 mL water per plant per day) and maintained at a relative humidity of 30%-40%, with a daytime temperature range of 21-27 °C, and a nighttime temperature range from 19-21 °C. Two-year-old walnut plants of the rootstock cultivar Vlach, produced by tissue culture,

were obtained from Sierra Gold Nursery (Yuba City, CA) and maintained in the EHS greenhouse at UC Davis under the same condition as described above for *D. stramonium*.

Pathogenicity test on *D. stramonium*

I used 45-day-old *D. stramonium* plants for the stab-inoculation method described previously [22]. Five plants per strain was inoculated and plants were distributed in a randomized complete block design (RCBD). The experiment was repeated three times and conducted from June to October 2019. The numbers of *D. stramonium* plants with tumors was recorded at 10 d, 13 d and 16 d post inoculation (dpi). The tumor fresh weight was recorded 50 dpi. The data of the three experiments were combined for final data analysis. All data were fitted in the model tumor_fresh_weight ~ Block + Isolate + (1|Position) to test for significant differences among the 30 strains. Furthermore, a regression analysis was performed as stated below.

Pathogenicity test on walnut rootstock Vlach

The 30 *A. tumefaciens* strains were inoculated into 2-year-old walnut seedling in two independent experiments performed on April 18th, 2019, and April 21st, 2019. Three plants were inoculated for each strain in each experiment. The walnut tumors were harvested and measured on June 9th, 2019 and June 12th, 2019 as described previously. The data of the two experiments were combined for final data analysis. Data analysis was performed as described below.

Antibiotics test

Eight classes of antibiotic were selected based on their mode of action: carbenicillin (inhibits cell wall synthesis), streptomycin (inhibits translation), vancomycin (inhibits cell wall synthesis), chloramphenicol (inhibits translation), rifampin (inhibits RNA synthesis), erythromycin (inhibits translation), tetracycline (inhibits translation), and ciprofloxacin (inhibits DNA synthesis). Furthermore, the antibiotics that inhibit translation all have different targets [23]. One hundred microliters of overnight culture adjusted to 10^7 cells/mL was spread on plates of TSBA medium. Immediately after, an antibiotic

impregnated disk (BD BBL™ Sensi-Disc™) was placed at the center of the plate. Then, the plate was incubated at 27 °C for 24 h. The zone of inhibition was measured for each strain and each strain was tested three times. The experiment was repeated three times. All data was analyzed as described below.

K84 bioassay test [24]

The Stonier and Kings B media were prepared following the previously described recipe [24]. I then streaked strain K84 on KB media and incubated it at 28 °C for 48-72 h. A single colony of K84 was selected, inoculated into 3 mL of AB liquid medium in a 10 mL tube, and grown at 28 °C for ~20 h in a shaker. Then 2 mL of the K84 culture was transferred to a 2 mL sterile Eppendorf tube and centrifuged at 20817 x g for 1 min. The supernatant was removed, and the pellet was resuspended in 1 mL sterile 0.9% NaCl and centrifuged at 20817 x g for 1 min. This process was repeated three times to remove the AB media. The pellet was suspended in 600 microliter DI water and the suspension was adjusted to optical density (OD_{600}) to 1.0, which equals ~10^9 cells/ml. I inoculated 5 microliters of the adjusted suspension on the center of the Stonier medium containing plate, making sure the Stonier medium surface was dry when doing this. The Stonier medium with K84 strain at the center was incubated at 28 °C for 48 h. Then, the Stonier medium was exposed to chloroform vapor for 10-15 min to kill K84 within a sealed glass/metal container in the fume hood. These plates were allowed to air dry for 5 min before they were ready for use in the bioassay.

Test strains preparation: A single colony of the test strain was inoculated into 2 mL of 10% TSB liquid medium and incubated for ~20 h at 28 °C. One mL of the culture was then transferred to a 1.5 Eppendorf tube and centrifuged at 20817 x g for 1 min. After removing the supernatant, the pellet was resuspended and washed with 1 mL 0.9% NaCl three times. The third resuspension was adjusted to an OD_{600} = 1.0, and then further diluted to ~ OD_{600} = 0.6, which was the final concentration used for the test strains in the K84 sensitivity test.

K84 bioassay: For the stock solutions of K84, one M monobasic dihydrogen phosphate (NaH_2PO_4), 1 M dibasic mono-hydrogen phosphate (K_2HPO_4) were prepared. Using these stock solutions, I prepared 500 mL of 20 mM potassium phosphate (K_3PO_4) at pH 7.0 following phosphate buffer instruction. A 0.7% agar (0.7 g bacto-agar in 100 mL of 20 mM potassium buffer) suspension was prepared and autoclaved for 25 min. The agar was then cooled to 50 °C. I then placed 2 mL 0.7% agar in a two mL Eppendorf tubes, added 100 microliters of the test strain, mixed well, and immediately overlayed this suspension on the surface of previously prepared Stonier medium. The plates were incubated at 28 °C for 48 to 72 h, after which the diameter of any halos was measured at 48 and 72 h two time points.

Motility test

The motility test was conducted using 10% TSBA containing 3% agar as described by McClean and Kluepfel [25]. The overnight culture of each strain was first adjusted to an optical density of 1.0. Then, five microliters of this culture were diluted at a ratio of 1:20 with sterile water, and placed at the center of the plate. Three experiments were prepared for each strain, and the diameter of the motility halo was measured after incubation for 24 h at 28 °C. This experiment was repeated three times.

Growth Rate test

I used two media types to examine growth rates, i.e., TSB rich and AB poor.. A 96-well micro-plate with each well contains 200 microliters of either TSB or AB was used for culturing. Each strain was replicated in three wells, each containing a 200 microliters suspension consisting of the bacterial cells to medium ratio of 1:100. After inoculation, the 96-well plates were covered with MicroAmp™ Optical Adhesive Film and incubated in a Synergy HTX Multi-Mode Reader at 27 C and shaken for 36 h at 200 rpm. Optical density values were recorded every 30 min at a wavelength of 600 nm. The growth rate experiment was repeated three times. The doubling time and growth rate were calculated using growthcurver [26], lubridate [27], tidyverse, dplyr [28], and tidyselect [29] R packages.

Statistical data analysis

Statistical data analysis was performed in three steps. First, I checked the data quality by removing 0s and fit data to a linear regression using lm() function in the stats package[30] or a mixed linear regression model using lmer() function of lme4 [31] package. Second, a diagnostic plot analysis was run to check if the selected model fit the data (Fig. 3.2; Fig. 3.3; Fig. 3.5; Fig. 3.6) using the ggplot [32] R package. If residuals normality or residuals homogeneity were not met, log transformation was applied to make the data meet data normal distribution and variance homogeneity assumptions (Fig. 3.4; Fig. 3.7 – Fig. 3.15). Third, Anova and multiple comparisons of the model were run using emmeans [33], rcompanion [34], car [35], multcomp [36] R packages. If the *p* value was < 0.05 ($P < 0.05$), it was considered that a statistically significant difference among strains in terms of the phenotypic trait analyzed. Based on the ANOVA result, I performed additional multiple comparison to confirm which strains are significantly different from one another for each trait. The codes for statistical analyses were attached as the html file.

Results

To examine the virulence of the 30 strains, including C58 as a reference, I inoculated *D. stramonium* and hybrid walnut seedlings (*J. hindsi* X *J. regia*) of the rootstock Vlach. Forty-five dpi, I measured tumor size (tumor fresh weight). On both host plant species, different strains generated significantly different *D. stramonium* tumor sizes ($F=4.09$; $P < .0001$) (Table 3.3) and walnut tumor sizes ($F=2.067$; $P < 0.05$) (Table 3.1). The estimated mean and standard error of *D. stramonium* tumor ranges from 0 g to 1.34 ± 0.19 g (Fig. 3.1; Table 3.4.) and the estimated mean and standard error of walnut tumor ranges from 0 g to 1.97 ± 0.38 g (Table 3.2). Interestingly, Kin002 was the only strain that consistently generated the smallest tumors in both species in both experiments, with an estimated mean of 0 g on *D. stramonium* and CIs between -0.8038 g and 0.01 g, and an estimated mean of 0 g on walnut and confidence intervals (CIs) between -0.97 g and 0.53 g, respectively. These results show all pathogenic strains on both species have a wide range of virulence levels.

To test antibiotic resistance of the 30 strains, I used the disk diffusion method with TSBA, with the disk impregnated with antibiotic placed at the center. The strains exhibited different levels of susceptibility to the antibiotics tested based on the size of the zone of inhibition (Fig. 3.16). Yub001 was the only strain that was resistant to carbenicillin. The remaining strains were sensitive to carbenicillin with varying degrees of sensitivity as defined by the size of the zone. Anova result showed that there was a significant difference in carbenicillin resistance (F=63.712; *P* <2.2e-16) (Table 3.5). Multiple comparison (Table 3.6) results showed that the estimated inhibition zone ranged from 1.7 cm with a 95% CI [1.5-1.8] for Tul002 to 3.4 cm of C58 at the CIs between 3.3 cm to 3.5 cm.

ANOVA showed that there was significant difference in chloramphenicol sensitivity of the 30 strains (F=78.337; *P*<2.2e-16) (Table 3.7). Multiple comparison (Table 3.8) demonstrated that the estimated inhibition zone ranged from 0.7 cm of Kin003 (CIs between 0.6 cm to 0.8 cm) to 2.5 cm of But001 (CIs between 2.4 cm to 2.6 cm). Thus, Sta004 and But001 were the two most sensitive strains to chloramphenicol, whereas Kin002 was the only strain resistant to chloramphenicol.

All 30 strains were sensitive to ciprofloxacin and erythromycin. ANOVA showed that there was a significant difference in the inhibition zones for both ciprofloxacin (F=34.739; *P* < 2.2e-16) (Table 3.9) and (F=18.306; P < 2.2e-16) (Table 3.11). The estimated inhibition zone of ciprofloxacin ranged from 1.9 cm for Kin002 (CIs: 1.76 cm - 2.04 cm) to 3.83 cm of C58 (CIs: 3.7 cm – 4.0 cm) (Table 3.10). Twenty-three out of thirty strains had an estimated inhibition zone > 3.0 cm. The estimated inhibition zone ranged from 0.7 cm of But002 (CIs: 0.6 cm - 0.8 cm) to 2.0 cm of Yub001 (CIs: 1.9 cm - 2.1 cm) (Table 3.12).

Sta003 was resistant to rifampin, whereas all other strains were sensitive to rifampin. ANOVA also showed that there was a significant difference in the size of the inhibition zone (F=51.96; *P* < 2.22e-16) (Table 3.13). The estimated inhibition zones of sensitive strains were similar ranging from 1.1 cm of Sta004 (CIs: 1.1 cm - 1.2 cm) to 1.6 cm of C58 (CIs: 1.5 cm - 1.7 cm) (Table 3.14).

Similar to ciprofloxacin and erythromycin, all 30 strains were sensitive to tetracycline. However, ANOVA demonstrated that there was a significant difference in the size of the zone of inhibition (F=39.445; *P* < 2.2e-16) (Table 3.15). The estimated inhibition zone ranged from 2.2 cm of Sta001 (CIs: 2.1 cm - 2.4 cm) to 3.70 cm of Yub001 (CIs: 3.6 cm - 3.8 cm) (Table 3.16.).

Nearly all 30 strains were resistant to streptomycin and vancomycin. Yub001 was the only exception, i.e., sensitive to streptomycin with an inhibition zone of 1.0 cm. Kin003 was the only strain sensitive to vancomycin with an inhibition zone of 0.6 cm (Fig. 3.16).

K84 sensitivity tests revealed that 15 strains were resistant to Agrocin84, and 15 strains were sensitive. ANOVA showed that there was a significant difference among the 15 sensitive strains (Table 3.17), with an inhibition zone ranging from 2.7 cm for Tul003 (CIs: 2.7 cm – 3.0 cm) to 5.7 cm for SJ001 (CIs: 5.4 cm – 5.7 cm) (Table 3.18).

I tested the strains for motility by growing them on a 10% TSBA minimal medium. The Anova result showed that there was a significant difference in the capability to move through semi-solid media (Table 3.19). The estimated mean of the diffusion zone ranged from 3.2 cm for CL001 (CIs: 3.1 cm - 3.3 cm) to 4.5 cm for Yub001 (CIs: 4.4 cm - 4.5 cm). This result shows that Yub001 is the most motile strain and CL001 is the least.

The doubling time of the 30 strains in both nutrient rich (TSB) and poor (AB) media was calculated. In general, most strains grew faster in TSB than in the AB medium. However, seven strains, including C58, CL001, Kin002, Sut001, Teh001, Tul001, and Yub002, grew similarly in both media. In TSB medium, the 30 strains exhibited a doubling time ranging from 1 h to 4 h. In AB media, the strains exhibited a doubling time ranging from 1.5 h to 10.5 h. Strain Yub001 grew fastest in the TSB with a median doubling time of 1.25 h, whereas in AB, it grew the slowest with a median doubling time of 9 h. The second fast growing strains in TSB were But001 and But002. Sta003 was the slowest growing strain in TSB. In contrast, interestingly, C58 and CC001 grew fastest in the nutrient poor AB medium with median doubling times <

2 h and also had a doubling time of ~2.0 h in TSB. Most strains grew more slowly in AB medium, including But001, But002, CL001, and Tul003 (Fig. 3.17). The growth rates of all 30 strains can be found in Fig. 3.18.

PCA analysis was performed on growth rates in TSB and AB media, *D. stramonium* and walnut tumor, K84 sensitivity, and motility (Fig. 3.19). Growth rates in AB and TSB media were negatively correlated. Growth rates on AB medium and walnut tumor was positively correlated. Motility was positively correlated with growth rate in TSB medium, whereas it was negatively correlated with growth rate in AB medium. Motility was not correlated with tumor size (Fig. 3.19).

Discussion

Agrobacterium tumefaciens Ti plasmids harbor the key genetic components required to induce tumor formation in susceptible hosts [37]. These genetic components, which mediate tumor formation include the transfer DNA (T-DNA) region, a ~20 kb component of the Ti plasmid, and a *vir* gene encoding type IV secretion system (T4SS), which is responsible for processing and transferring T-DNA from pathogen to host cells [37]. In pathogenicity tests of 30 strains, three strains (But002, Tul002, SJ003) were not pathogenic, i.e., failed to induce tumors on both *D. stramonium* and walnuts. Genomic data support this conclusion and will be presented in Chapter 4. Even though these strains were isolated from young walnut tumor tissue, they were found to be devoid of a Ti-plasmid. This suggests a mixed population of pathogenic and non-pathogenic strains in tumors. The remaining 26 virulent strains and C58 exhibited significant variability in virulence on *D. stramonium* and walnut.

Tumor development is considered to be one strategy used by pathogenic *A. tumefaciens* strains to obtain energy and nutrient from host plants [38]. Secreting antibiotics is also a common strategy for bacteria to compete and survive in an adverse environment. Among the eight antibiotics evaluated in the present study, tetracycline, erythromycin, and ciprofloxacin were three antibiotics, to which all strains exhibited various levels of sensitivity. Streptomycin is broadly used in the management of fire blight disease in apple and pear orchards [39]. Kasugamycin is an aminoglycoside antibiotic, which is effective in

controlling walnut blight when used in rotation with copper (Cu) alone or with mancozeb [40]. The rotation materials have different modes of action in targeting bacteria, which helps to limit development of antibiotic resistance in the pathogen.

Horizontal gene transfer is one mechanism that contributes to antibiotic resistance genes jumping from one bacterial cell to another, leading to the antibiotic resistance spreading throughout a population [41]. Though streptomycin is mainly used in agriculture, resistance for this antibiotic has been identified not only in human, animal and plant pathogens, but also in a wide range of environmental bacteria [17]. This global scale discovery of streptomycin resistance may have a significant impact on human health. Thus, HGT under selective pressure due to use of streptomycin in agriculture operations may explain the high incidence of streptomycin resistance detected in *A. tumefaciens*. This suggests the limited use of antibiotics for crown gall control.

Tetracycline is more than a clinical drug. In nature, it is a broad-spectrum polyketide and exhibits antimicrobial activity against Gram-negative and Gram-positive bacteria, spirochetes, obligate intracellular bacteria, and protozoan parasites [42, 43]. Tetracycline resistance has been prevalent in clinical environments, but is not yet a major concern in agriculture [17]. My assessment of tetracycline resistance of *A. tumefaciens* strains also confirmed this notion. Thus, thirty test strains were sensitive to tetracycline. Given the prevalence of antibiotic resistance to tetracycline in clinical settings, application of tetracycline for crown gall management should be limited even though all strains were susceptible.

Most strains were sensitive to rifampin, carbenicillin, and chloramphenicol with the following exceptions. Yub001 was resistant to carbenicillin, Sta003 resistant to rifampin, and Kin002 resistant to chloramphenicol. The remaining strains exhibited variable sensitivity to rifampin, carbenicillin, and chloramphenicol. Heavy antibiotic use combined with the HGT of antibiotic resistance genes may enhance frequency of antibiotic resistance in *A. tumefaciens*. If antibiotic usage is not properly managed, the

overuse of antibiotics may place selective pressure to *A. tumefaciens* strains, and the frequency of antibiotic resistance will continue to increase.

The antibiotic resistance profile of the strains also shows that for each antibiotic, strains from different counties exhibited similar antibiotics resistance profile even though these counties are geographically distant. Therefore, development of these antibiotic resistance profiles are likely independent events, indicating these strains may be under a common selective factor in the environment. In addition, because the five largest nurseries in California provide walnut clonal or seedlings rootstock to growers across the entire central valley a similar antibiotic resistance profile may be expected if the propagation material was infected with *A. tumefaciens*. Also, since a given grower normally buys rootstocks from more than one nursery, this also may contribute to strains from different counties exhibiting similar antibiotic resistance profiles. Finally, the movement of machinery with contaminated soil between fields may also contribute to the dissemination of this pathogen.

Some bacteria that release antibiotics have been exploited in the development of commercial products for the biological control of plant disease. *Agrobacterium* biocontrol strain K84 is one such strain that produces a nucleotide bacteriocin, agrocin84, which can inhibit RNA and DNA synthesis and is a mimic of agrocinopine. *A. tumefaciens* strains capable of metabolizing agrocinopine are sensitive to agrocin84 [20]. Controlling crown gall using K84 has been shown effective in numerous production systems [44]. However, some *Agrobacterium* strains isolated from rootstocks of stone fruit trees were resistant to agrocinK84 [45]. The K84 sensitivity bioassay revealed that 15 out of 30 *A. tumefaciens* strains were resistant to the K84 strain. These results indicate that biocontrol products using strain K84 would likely not be effective for crown gall management in walnut orchards in CA. More effective biocontrol strains are required for crown gall management.

The evolutionary trade-off between various traits, for example, the trade-off between growth rates and antibiotic resistance, the trade-off between antibiotic resistance and fitness in the absence of

the antibiotic, has been studied in microbiology [46]. The negative correlation between growth rate and resistance to fluoroquinolone and cephalosporin antibiotics was reported in *E. coli* [46]. The PCA results showed a negative correlation between growth rate in AB and rifampin resistance, AB growth rate and erythromycin resistance, and TSB growth rate and carbenicillin resistance. However, positive correlations were observed between AB growth rate and carbenicillin resistance, TSB growth rate and erythromycin and rifampin resistance. These relationships may evolve under varying environmental conditions [46]. A study examining the evolution of virulence *A. tumefaciens* showed that the host plant could favor growth of more nonpathogenic mutants [47]. In our experience, when attempting to isolate *A. tumefaciens* strains from walnut tumors or soil samples, more nonpathogenic *A. tumefaciens* strains tend to be detected. The PCA results also indicated that *A. tumefaciens* can be an opportunistic pathogen and that under nutrient limiting conditions *A. tumefaciens* tends to be more virulent in an effort to obtain nutrients from the host plant. However, *A. tumefaciens* strains growing in/on healthy tumor tissue most likely are not under limited nutrient conditions, which may result in a greater population of nonpathogenic derivatives.

My study provides preliminary data for understanding the evolutionary fitness and adaptation of *A. tumefacien* existing under commercial walnut orchard conditions in Central California. In summary, these results provide information, which may facilitate development of more effective crown gall management strategies including biocontrol strains, changing their growth environment, and host-controlled crown gall resistance.

Figures and tables

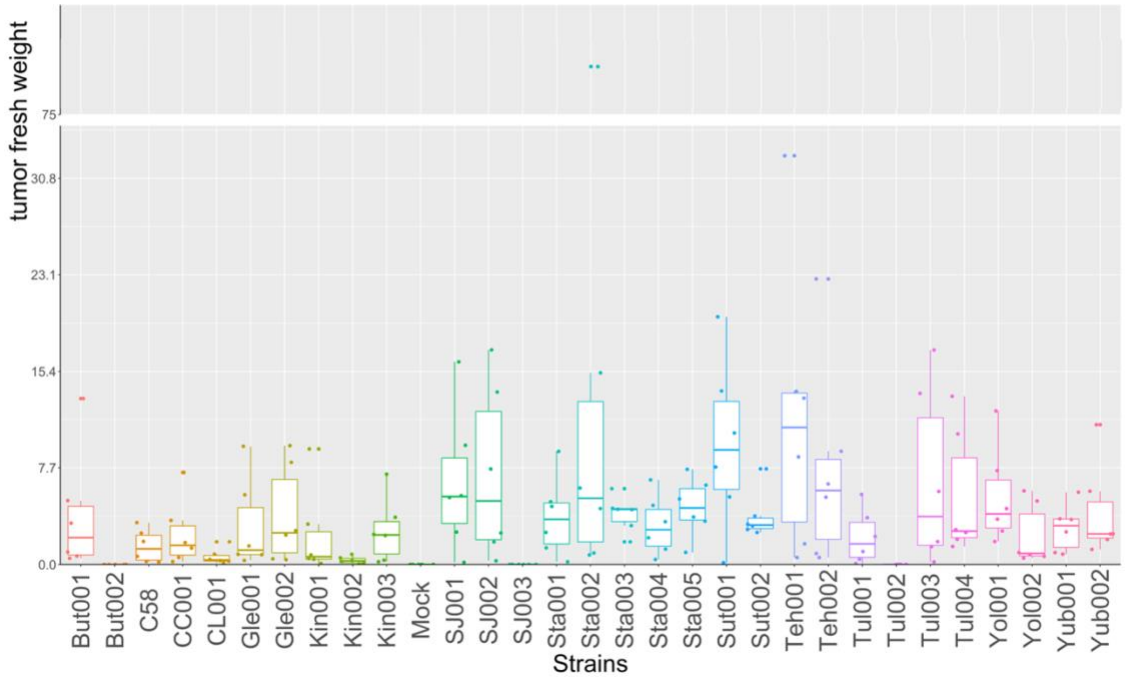


Fig. 3.1. Boxplot of fresh tumor weights on walnut rootstock Vlach generated by 30 *Agrobacterium* strains. Strains are ordered alphabetically. The x-axis indicates the 30 *Agrobacterium tumefaciens* strains examined. The y-axis is the fresh tumor weight in gram (g).

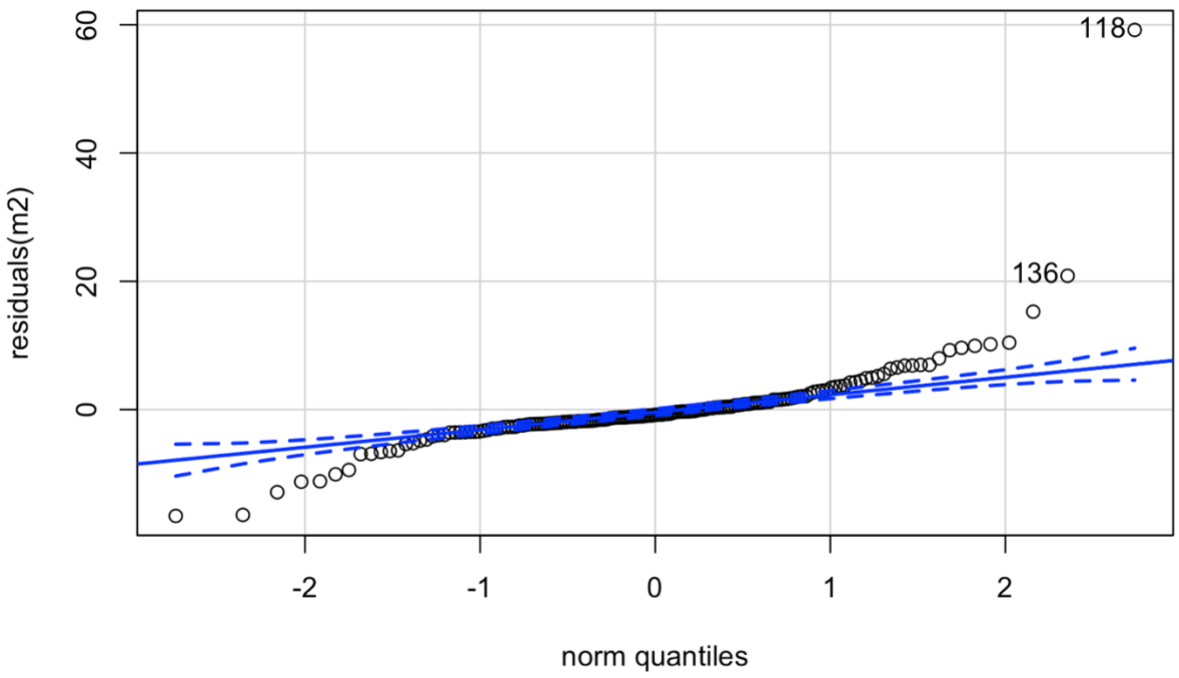


Fig. 3.2. The normal quantile plot of the residuals of the model for virulence on walnut rootstock. Open circle data points outside of the dashed line ranges indicates that the residuals were not normally distributed.

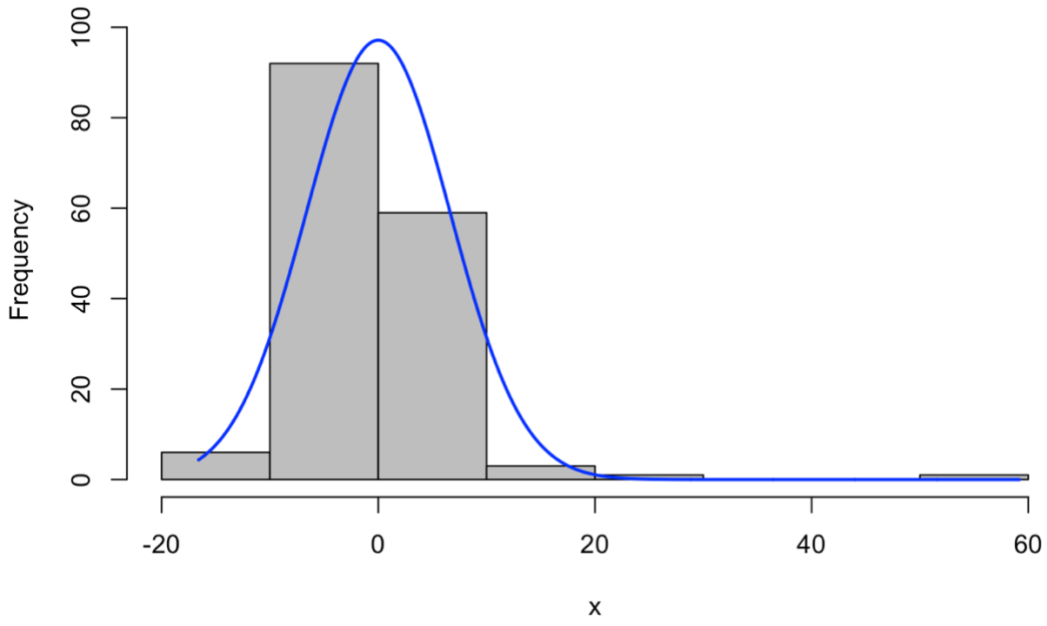


Fig. 3.3. Histogram of walnut tumor weight frequency distribution. The X-axis illustrates the classes of walnut tumor fresh weight. The Y-axis is the frequency of each fresh weight class. The bell-shaped curve indicates that the histogram is skewed to the left.

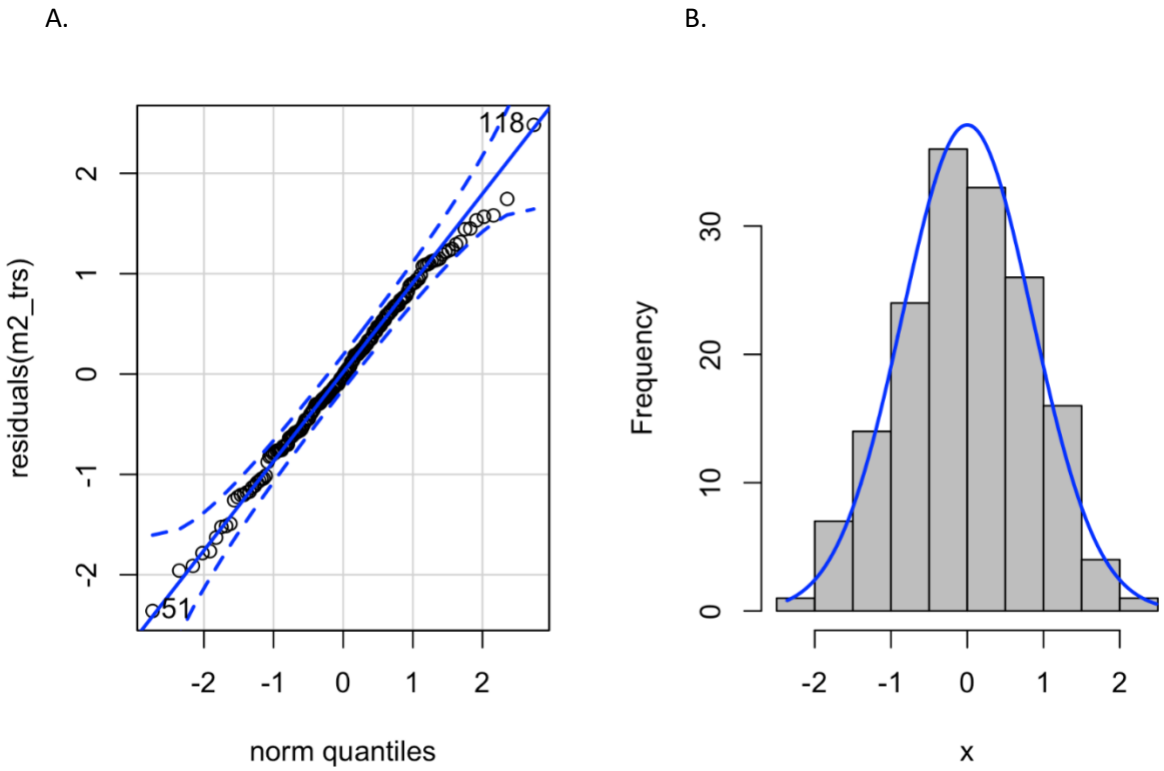


Fig. 3.4. The normal quantile plot of the residuals (A) and histogram of fresh weight frequency distribution (B) after the model is log transformed. Log transformation makes both residuals and fresh weight data very close to being normally distributed. The bell-shaped curve indicates the normal distribution after transformation.

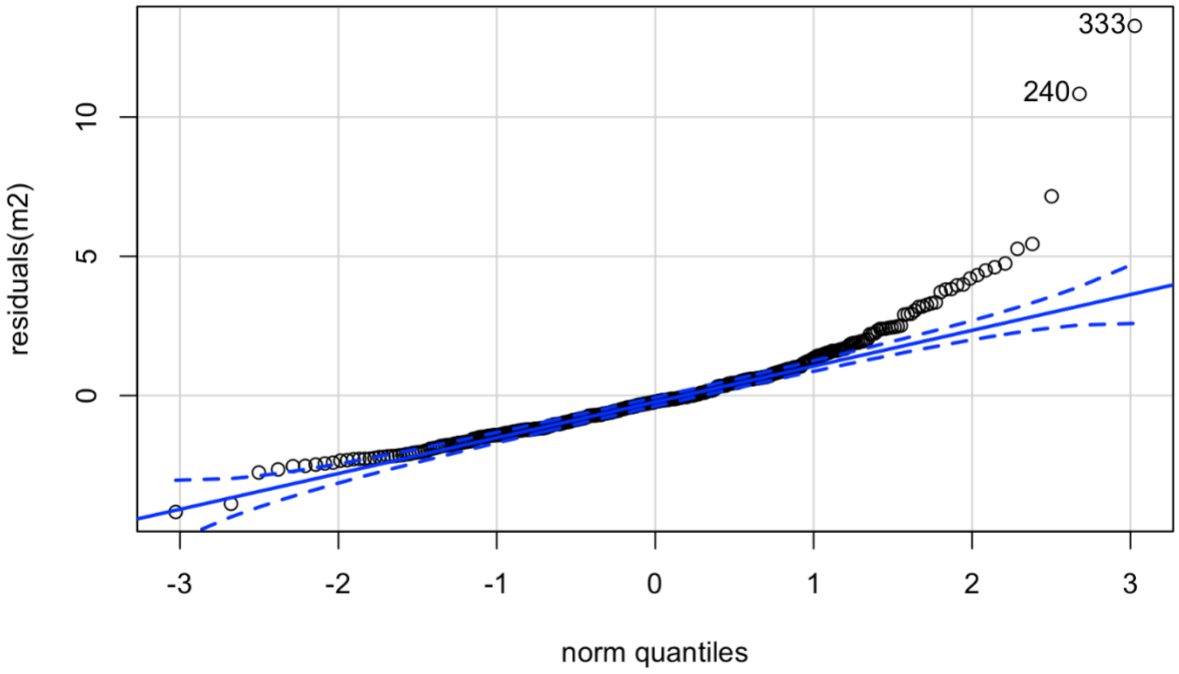


Fig. 3.5. The normal quantile plot of the residuals of the model for virulence on *Datura*. The X-axis is norm quantiles. The Y-axis is residuals(m2). Data points out of the dashed lines ranges indicated that the residuals were not normally distributed. The two dots marked as 240 and 333 are outliers.

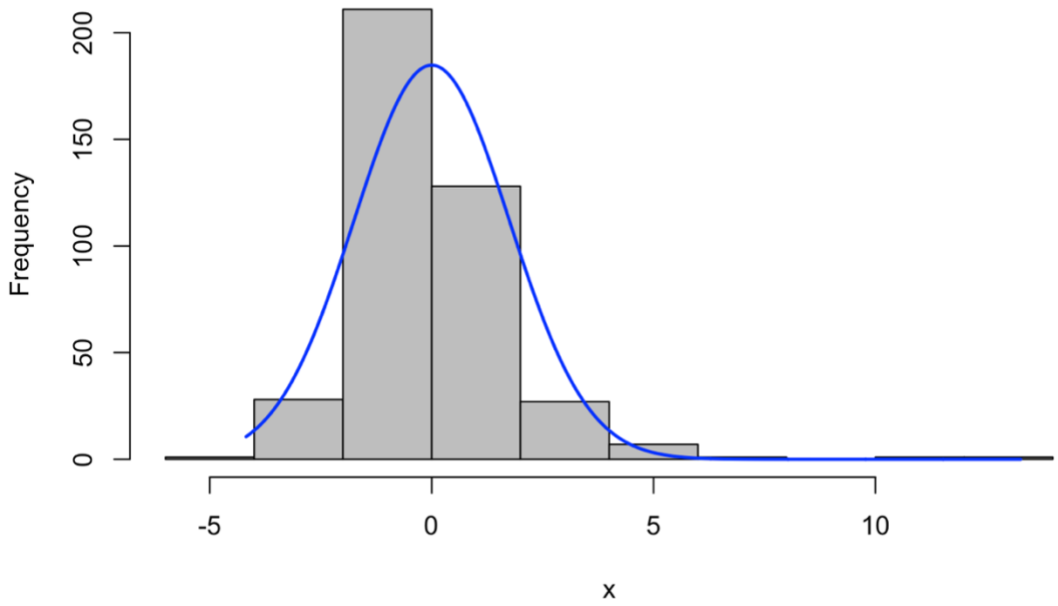


Fig. 3.6. Histogram of datura tumor weight frequency distribution. The X-axis is the classes of *Datura* tumor fresh weight. The Y-axis is the frequency of each fresh weight class. The bell-shaped curve showed that the histogram is skewed to the left.

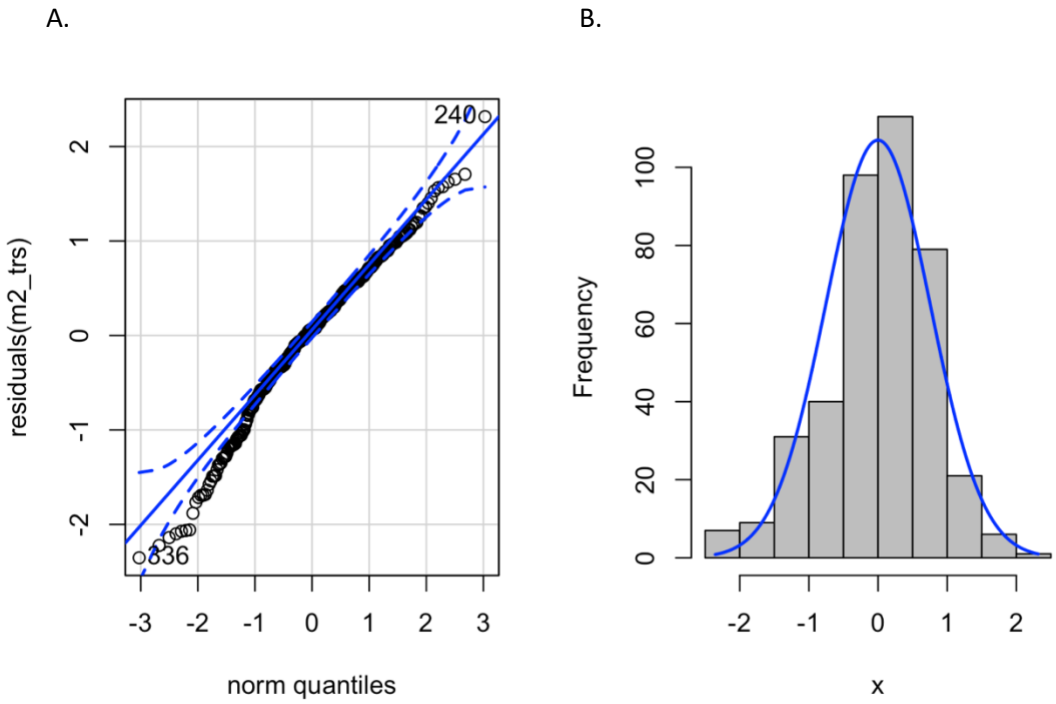


Fig. 3.7. The normal quantile plot of the residuals (left) and histogram of fresh weight frequency distribution (right) after the model is log transformed. A log transformation makes both residuals and *Datura* tumor fresh weight data be close to normal distribution, as the bell-shaped curve displayed.

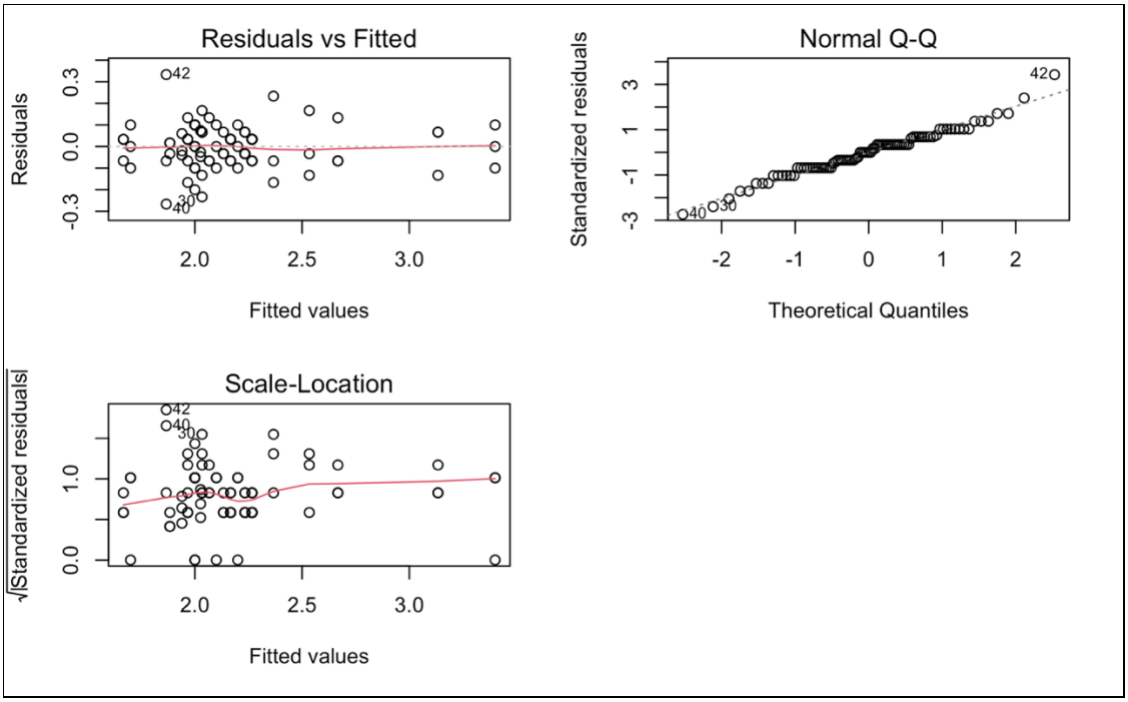


Fig. 3.8. Diagnostic plots for linear regression analysis of A. tumefaciens carbenicillin resistance. Residuals vs Fitted plot shows residuals do not have non-linear patterns. Normal Q-Q plot shows residuals are normally distributed as most residuals follow a straight line well. Scale-Location plot shows that the data meets the assumption of equal variance (homoscedasticity) as we see a horizontal line with equally (randomly) spread points.

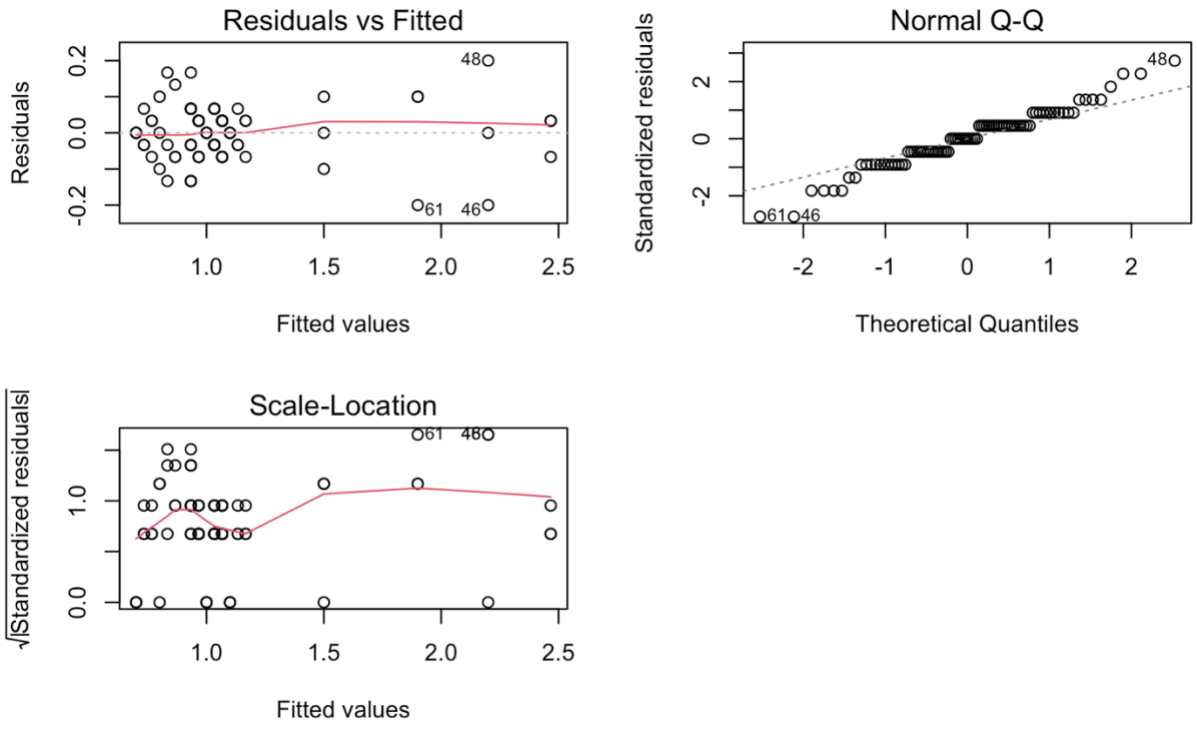


Fig. 3.9. Diagnostic plots for linear regression analysis of *A. tumefaciens* chloramphenicol resistance. Residuals vs Fitted plot shows residuals do not have non-linear patterns. Normal Q-Q plot shows residuals are to close normally distributed as most residuals follow a straight line well, thought data in both ends deviate a little bit. Scale-Location plot shows that the data meets the assumption of equal variance (homoscedasticity) as we see a horizontal line with equally (randomly) spread points.

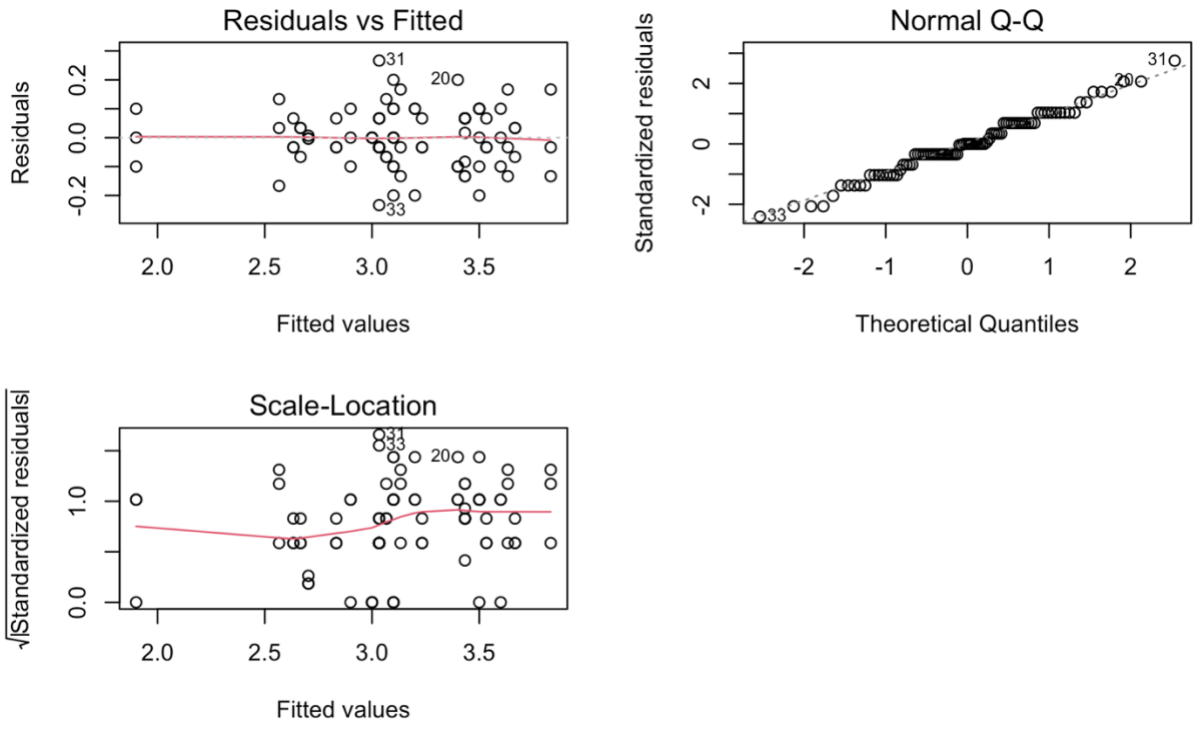


Fig. 3.10. Diagnostic plots for linear regression analysis of *A. tumefaciens* ciprofloxacin resistance. Residuals vs Fitted plot shows residuals do not have non-linear patterns. Normal Q-Q plot shows residuals are normally distributed as most residuals follow a straight line well. Scale-Location plot shows that the data meets the assumption of equal variance (homoscedasticity) as we see a horizontal line with equally (randomly) spread points.

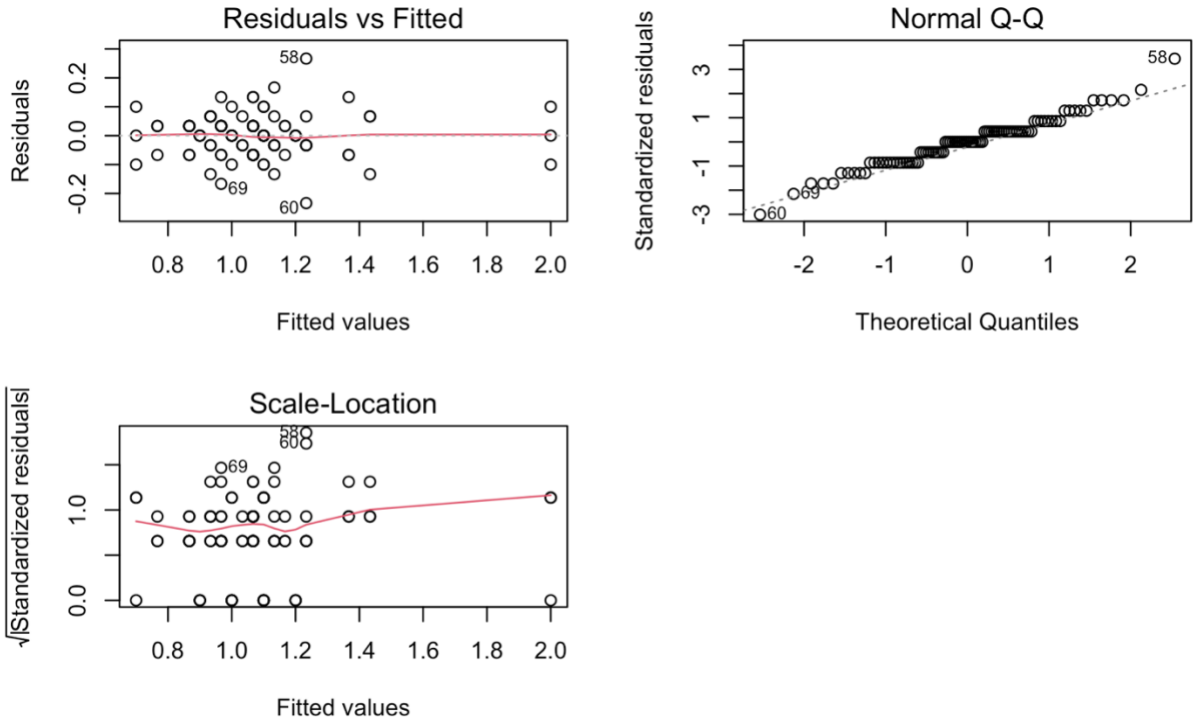


Fig. 3.11. Diagnostic plots for linear regression analysis of *A. tumefaciens* erythromycin resistance. Residuals vs Fitted plot shows residuals do not have non-linear patterns. Normal Q-Q plot shows residuals are normally distributed as most residuals follow a straight line well. Scale-Location plot shows that the data meets the assumption of equal variance (homoscedasticity) as we see a horizontal line with equally (randomly) spread points.

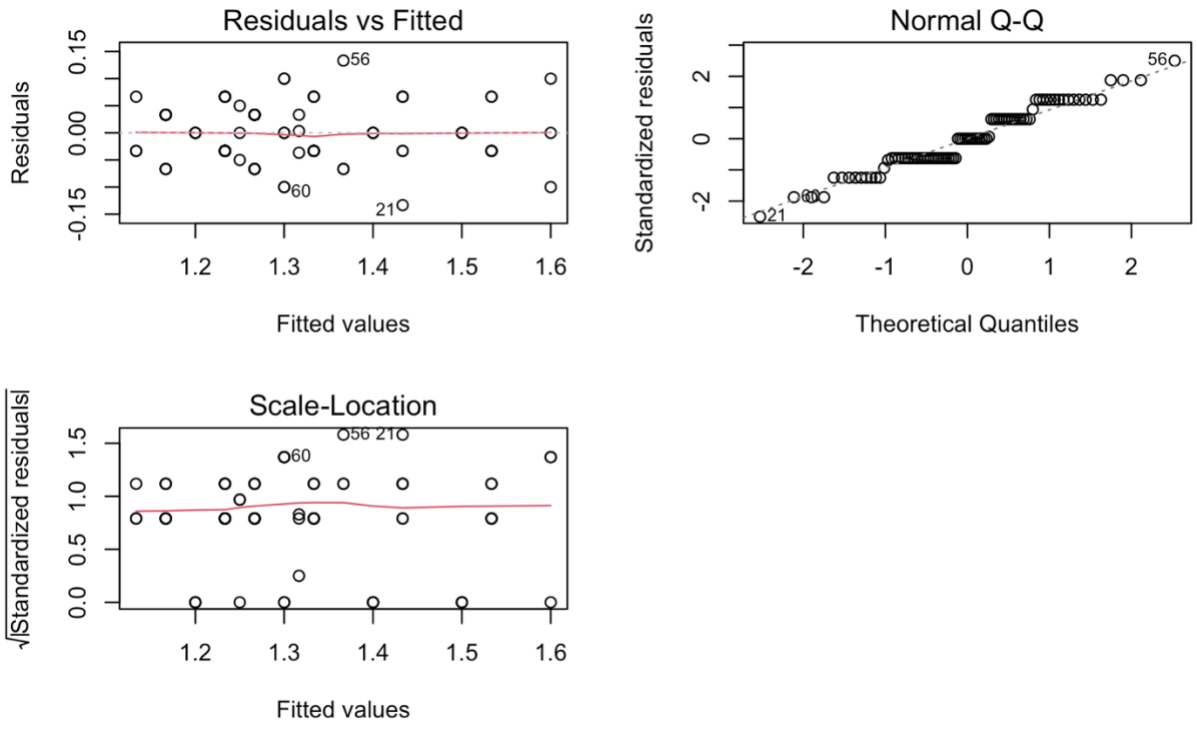


Fig. 3.12. Diagnostic plots for linear regression analysis of *A. tumefaciens* rifampin resistance. Residuals vs Fitted plot shows residuals do not have non-linear patterns. Normal Q-Q plot shows residuals are normally distributed as most residuals follow a straight line well. Scale-Location plot shows that the data meets the assumption of equal variance (homoscedasticity) as we see a horizontal line with equally (randomly) spread points.

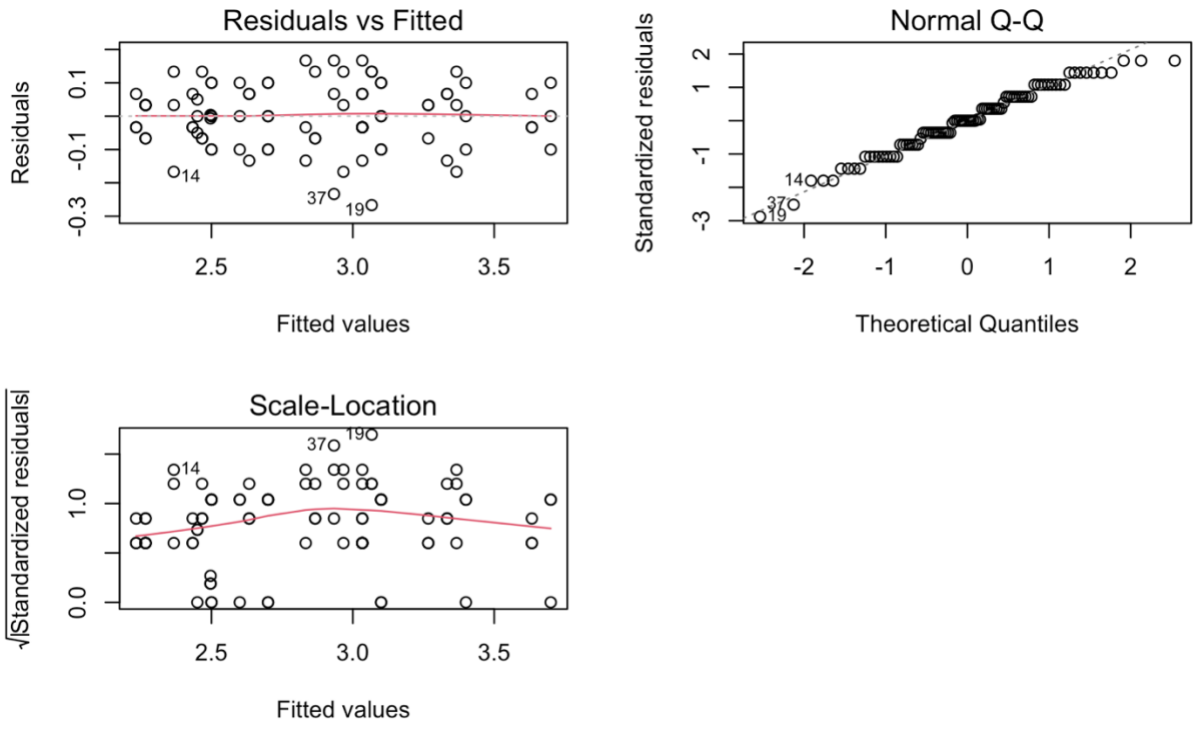


Fig. 3.13. Diagnostic plots for linear regression analysis of *A. tumefaciens* tetracycline resistance. Residuals vs Fitted plot shows residuals do not have non-linear patterns. Normal Q-Q plot shows residuals are normally distributed as most residuals follow a straight line well. Scale-Location plot shows that the data meets the assumption of equal variance (homoscedasticity) as we see a horizontal line with equally (randomly) spread points.

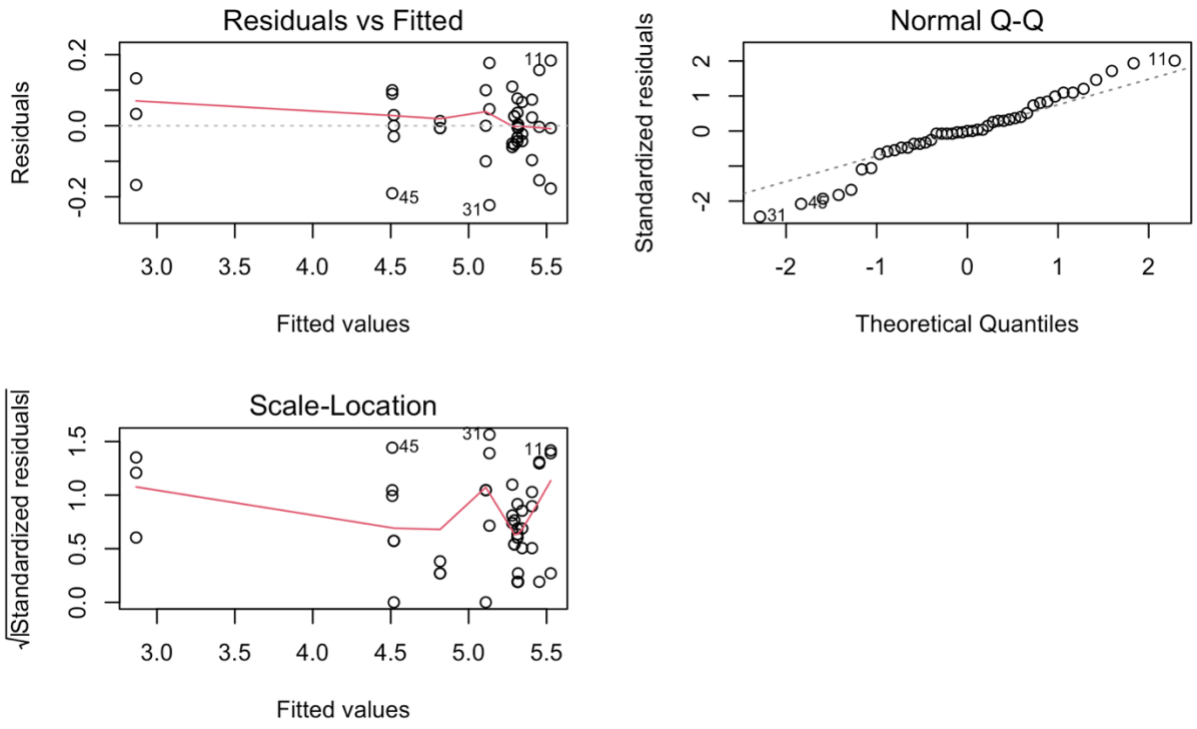


Fig. 3.14. Diagnostic plots for linear regression analysis of *A. tumefaciens* agrocin-84 sensitivity. Residuals vs Fitted plot shows residuals do not have non-linear patterns. Normal Q-Q plot shows residuals are close to normally distributed, though data in both ends slightly deviate. Scale-Location plot shows that the data meets the assumption of equal variance (homoscedasticity) as we see a horizontal line with equally (randomly) spread points.

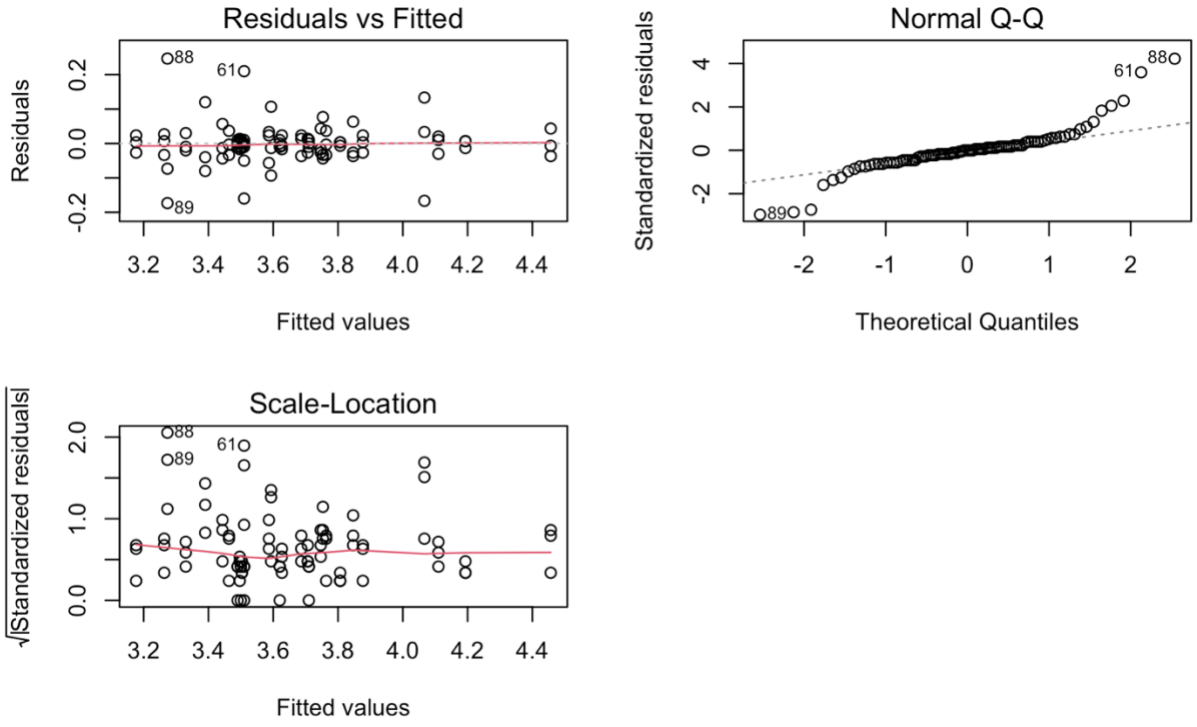


Fig. 3.15. Diagnostic plots for linear regression analysis of *A. tumefaciens* motility. Residuals vs Fitted plot shows residuals do not have non-linear patterns. Normal Q-Q plot shows residuals are close to normally distributed, though data in both ends slightly deviate. It is acceptable to continue the data analysis. Scale-Location plot shows that the data meets the assumption of equal variance (homoscedasticity) as we see a horizontal line with equally (randomly) spread points.

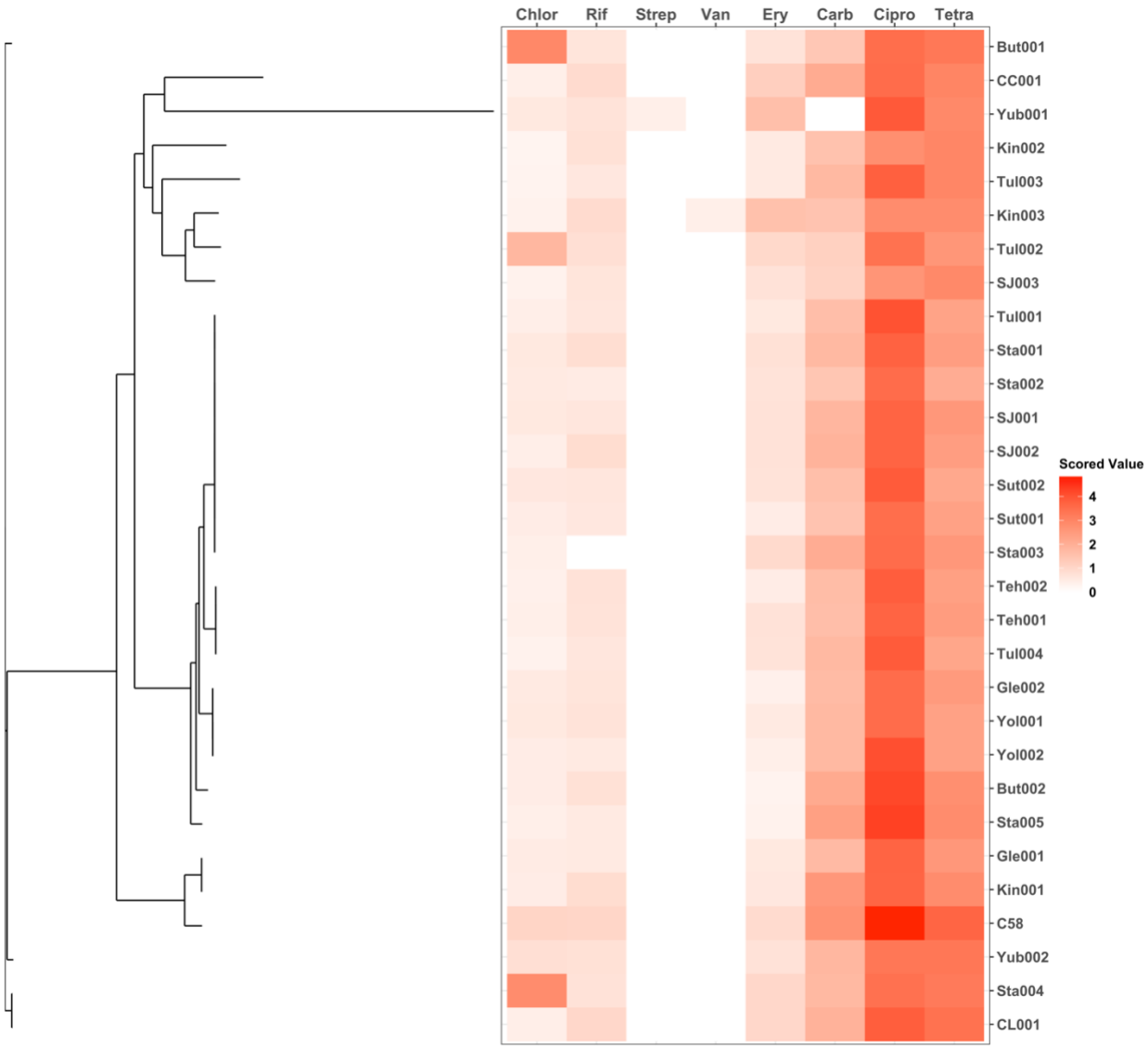


Fig. 3.16. Resistance profiles of 30 strains to eight antibiotics. The tree on the left is the dendrogram of 30 *A. tumefaciens* strains which are indicated vertically on the right. The profile on the right represents 8 antibiotics resistance data. The color changing from white gradually to red indicates changes from resistant (white color) to sensitive (red color).

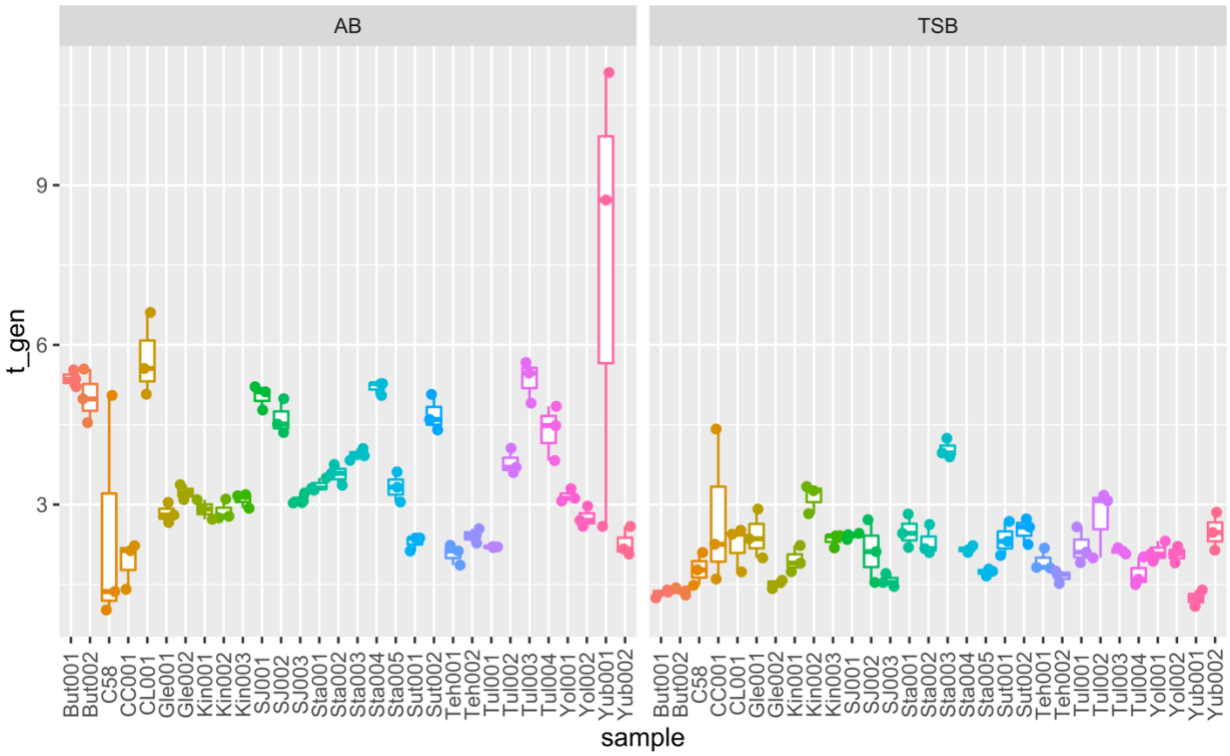


Fig. 3.17. Boxplot of doubling time (hours) of 30 *Agrobacterium* strains growing in AB and TSB medium. The X-axis indicates the individual strains. The Y-axis is doubling time (hours). In TSB medium, But001, But002, and Yub001 have the smallest doubling time of less than 1.5 hours. Sta003 has the greatest doubling time of greater than four hours. CC001 has the largest variance. In AB medium, C58 and CC001 have the smallest doubling time around 2.5 hours. Yub001 has the greatest doubling time of greater than six hours and largest variance.

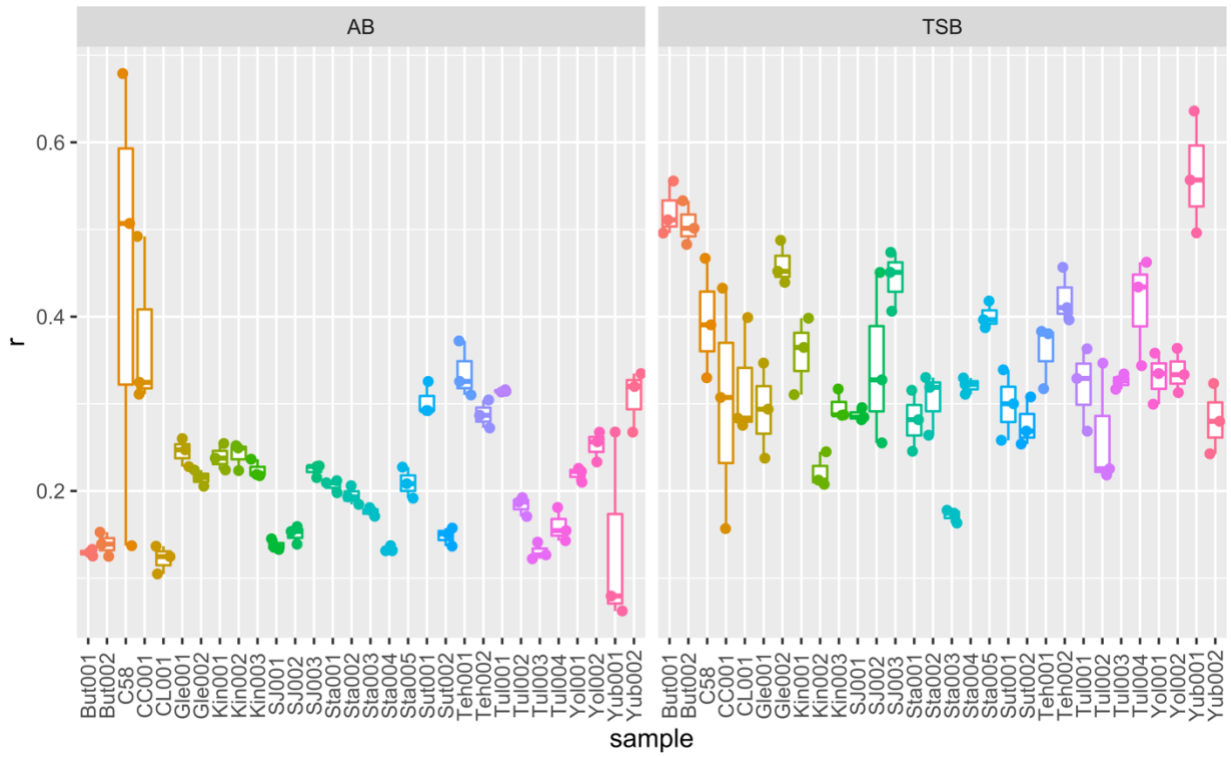
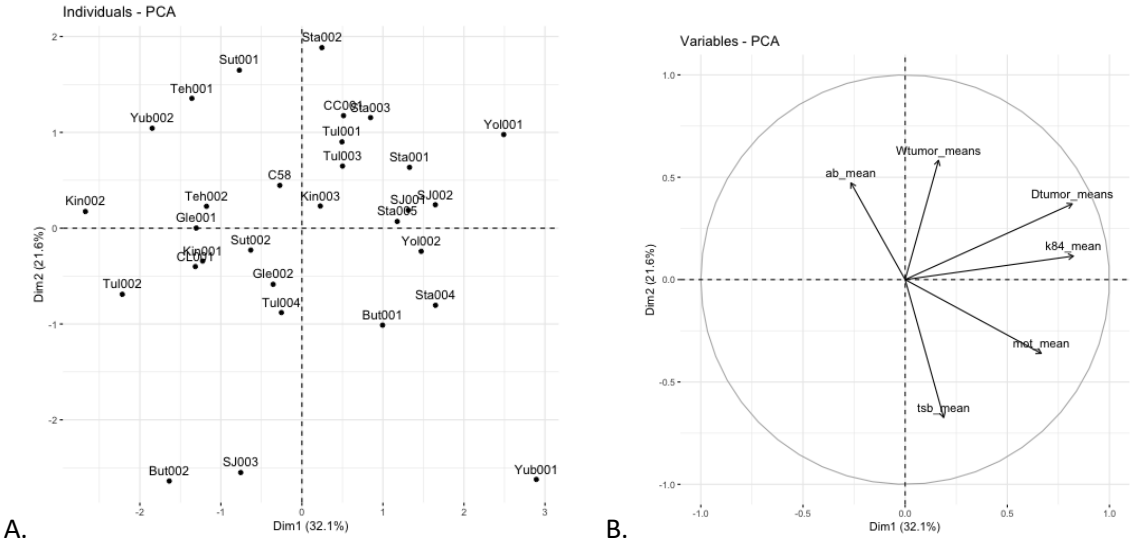


Fig. 3.18. Growth rate (minutes^{-1}) of 30 *A. tumefaciens* strains growing in AB and TSB medium. The X-axis is the strains name. The Y-axis is the growth rate. In TSB medium, But001, But002, and Yub001 have the highest growth rate at greater than 0.5, while Sta003 has the lowest growth rate around 0.17. In AB medium, C58 and CC001 have the highest growth rate, greater than 0.3, while Yub001 has the lowest growth rate at less than 0.1.



A. B.

Fig 3.19. Two-dimensional Principal components analyses (PCA) of phenotypic traits. A (left) is the individual plot. B (right) is the variables correlation plot. Growth rate in TSB rich medium and in AB poor medium are negatively correlated. Growth rate in AB poor medium and walnut tumor size are positively correlated.

Table 3.1. Analysis of variance of walnut tumor weight. Column 1 is degrees of freedom(Df). Column 2 is Sum of squares (Sum-Sq). Column 3 is Mean Squares (Mean-Sq). Column 4 is F ratio. Column 5 is p value.

Response: log(weight + 0.5)

	1-Df	2-Sum Sq	3-Mean Sq	4-F value	5-Pr(>F)
strains	26	46.61514	1.79289	2.06661	0.004045
Residuals	135	117.1194	0.867551		

Signif. codes: '**' 0.01

Table 3.2. Multi-comparison of walnut fresh tumor weight generated by 27 *A. tumefaciens* strains. Column 1 is the strain name. Column 2 is the estimated mean of the tumor fresh weight of each strain based on the model. Column 3 is standard error. Column 4 is degrees of freedom. Columns 5 and 6 are the lower confidence level and upper confidence level respectively. Column 7 represents the statistical group in terms of significance. Confidence level = 0.95. p-value adjustment: Tukey method for comparing a family of 27 estimates. Significance level = 0.05.

1-strains	2-Emmean	3-SE	4-Df	5-lower.CL	6-upper.CL	7-.group
Kin002	-0.22	0.38	135	-0.97	0.53	a
CL001	0.01	0.38	135	-0.74	0.76	ab
C58	0.47	0.38	135	-0.29	1.22	ab
Kin001	0.53	0.38	135	-0.22	1.28	ab
Tul001	0.68	0.38	135	-0.08	1.43	ab
Yol002	0.73	0.38	135	-0.02	1.48	ab
CC001	0.77	0.38	135	0.01	1.52	ab
Kin003	0.85	0.38	135	0.1	1.6	ab
Gle001	0.85	0.38	135	0.1	1.6	ab
But001	1.04	0.38	135	0.29	1.79	ab
Sta004	1.07	0.38	135	0.32	1.82	ab
Yub001	1.07	0.38	135	0.32	1.83	ab
Gle002	1.08	0.38	135	0.33	1.83	ab
Sta001	1.16	0.38	135	0.41	1.92	ab
Yub002	1.32	0.38	135	0.56	2.07	ab
Sut002	1.41	0.38	135	0.66	2.16	ab
Tul003	1.42	0.38	135	0.67	2.17	ab
Sta003	1.46	0.38	135	0.71	2.21	ab
Tul004	1.48	0.38	135	0.73	2.24	ab
Sta005	1.49	0.38	135	0.74	2.25	ab
SJ002	1.55	0.38	135	0.8	2.31	ab
SJ001	1.57	0.38	135	0.82	2.32	ab
Teh002	1.58	0.38	135	0.83	2.33	ab
Yol001	1.61	0.38	135	0.85	2.36	ab
Sta002	1.86	0.38	135	1.11	2.61	b
Sut001	1.92	0.38	135	1.17	2.67	b
Teh001	1.97	0.38	135	1.22	2.72	b

Results are given on the log(mu + 0.5) (not the response) scale.
Confidence level used: 0.95
Results are given on the log (not the response) scale.
P value adjustment: tukey method for comparing a family of 27 estimates
Significance level used: alpha = 0.05

Table 3.3. Analysis of variance of 27 pathogenic strains using lme model. Column 1 is the Chisq value; column 2 is the freedom of numerator; column 3 is the p-value.

	1-Chisq	2-Df	3-Pr(>Chisq)
strains	106.2916	26	1.13E-11
block	0.025194	1	0.873885

Table 3.4. Multi-comparison of datura fresh tumor weight generated by 27 *A. tumefaciens* strains. Column 1 is the strain name. Column 2 is the estimated mean of the tumor fresh weight of each strain based on the model. Column 3 is standard error. Column 4 is degrees of freedom. Columns 5 and 6 are the lower confidence level and upper confidence level respectively. Column 7 represents the statistical group in terms of significance. Confidence level = 0.95. p-value adjustment: Tukey method for comparing a family of 27 estimates. Significance level = 0.05.

1-strains	2-emmean	3-SE	4-df	5-lower.CL	6-upper.CL	7-.group
Kin002	-0.39	0.21	75.447	-0.8	0.01	a
Teh001	0.27	0.2	91.87604	-0.13	0.67	ab
C58	0.43	0.18	159.4682	0.06	0.79	abc
Gle001	0.43	0.19	137.1998	0.07	0.8	abc
Yub002	0.44	0.19	114.6926	0.07	0.82	abc
Teh002	0.49	0.18	212.1309	0.13	0.84	abc
Kin001	0.49	0.18	158.5338	0.13	0.86	abc
Kin003	0.51	0.18	162.1278	0.15	0.87	abc
Sut002	0.54	0.19	130.2805	0.18	0.91	abc
Tul004	0.73	0.19	108.642	0.36	1.1	bc
Sut001	0.74	0.18	165.6285	0.38	1.1	bc
CC001	0.78	0.19	138.497	0.42	1.15	bc
Sta002	0.82	0.18	200.1242	0.47	1.18	bc
SJ001	0.83	0.18	169.9166	0.47	1.19	bc
Gle002	0.85	0.18	159.3375	0.49	1.21	bc
CL001	0.86	0.2	84.94108	0.46	1.27	bc
Tul001	0.9	0.19	116.7349	0.53	1.27	bc
But001	0.9	0.18	165.8907	0.54	1.26	bc
Sta004	0.93	0.18	164.8889	0.57	1.29	bc
SJ002	0.97	0.18	191.7754	0.61	1.32	bc
Yol002	1	0.18	167.5937	0.64	1.36	bc
Sta003	1.03	0.18	142.6097	0.66	1.39	bc
Sta001	1.05	0.18	160.6969	0.68	1.41	bc
Yub001	1.07	0.18	162.7325	0.7	1.43	bc
Sta005	1.25	0.18	169.5977	0.88	1.61	bc
Tul003	1.32	0.19	124.0897	0.96	1.69	c
Yol001	1.34	0.19	116.6975	0.97	1.71	c

Degrees-of-freedom method: Kenward-Roger
Results are given on the log(mu + 0.5) (not the response) scale.
Confidence level used: 0.95
Results are given on the log (not the response) scale.
P value adjustment: Tukey method for comparing a family of 27 estimates
Significance level used: alpha = 0.05

Table 3.5. Analysis of variance of carbenicillin resistance of 30 A. tumefaciens strains. Column 1 is degrees of freedom (Df). Column 2 is Sum of squares (Sum-Sq). Column 3 is Mean Squares (Mean-Sq). Column 4 is F ratio. Column 5 is p value.

	1-Df	2-Sum Sq	3-Mean Sq	4-F value	5-Pr(>F)
strains	28	11.50949	0.411053	28.99669	8.43E-25
Residuals	58	0.8222	0.014176		

Signif. codes: '***' 0.001

Table 3.6. Multi-comparison of carbenicillin resistance of 30 *A. tumefaciens* strains. Column 1 is the strain name. Column 2 is the estimated mean of the inhibition zone of each train based on the model. Column 3 is standard error. Column 4 is degrees of freedom. Columns 5 and 6 are the lower confidence level and upper confidence level respectively. Column 7 represents the statistical group in terms of significance. Confidence level = 0.95. p-value adjustment: Tukey method for comparing a family of 30 estimates. Significance level = 0.05.

1-strains	2-emmean	3-SE	4-df	5-lower.CL	6-upper.CL	7-.group
Tul002	1.67	0.07	58	1.53	1.8	a
Kin002	1.7	0.07	58	1.56	1.84	ab
Sta001	1.87	0.07	58	1.73	2	abc
SJ003	1.88	0.07	58	1.75	2.02	abc
SJ002	1.94	0.07	58	1.8	2.08	abcd
Kin003	1.97	0.07	58	1.83	2.1	abcd
Sut002	1.97	0.07	58	1.83	2.1	abcd
Gle001	2	0.07	58	1.86	2.14	abcde
Tul004	2	0.07	58	1.86	2.14	abcde
Kin001	2	0.07	58	1.86	2.14	abcde
SJ001	2.03	0.07	58	1.89	2.16	abcde
But001	2.03	0.07	58	1.9	2.17	abcde
Sut001	2.03	0.07	58	1.9	2.17	abcde
Tul003	2.07	0.07	58	1.93	2.2	bcde
CC001	2.1	0.07	58	1.96	2.24	cde
Yub002	2.13	0.07	58	2	2.27	cde
Gle002	2.17	0.07	58	2.03	2.3	cdef
CL001	2.17	0.07	58	2.03	2.3	cdef
Sta002	2.2	0.07	58	2.06	2.34	cdef
Sta004	2.23	0.07	58	2.1	2.37	cdef
Yol001	2.23	0.07	58	2.1	2.37	cdef
Teh002	2.27	0.07	58	2.13	2.4	def
Tul001	2.27	0.07	58	2.13	2.4	def
Yol002	2.27	0.07	58	2.13	2.4	def
Teh001	2.37	0.07	58	2.23	2.5	efg
But002	2.53	0.07	58	2.4	2.67	fg
Sta003	2.67	0.07	58	2.53	2.8	g
Sta005	3.13	0.07	58	3	3.27	h
C58	3.4	0.07	58	3.26	3.54	h

Confidence level used: 0.95
P value adjustment: Tukey method for comparing a family of 30 estimates
Significance level used: alpha = 0.05

Table 3.7. Analysis of variance of chloramphenicol resistance of 30 *A. tumefaciens* strains. Column 1 is degrees of freedom (Df). Column 2 is Sum of squares (Sum-Sq). Column 3 is Mean Squares (Mean-Sq). Column 4 is F ratio. Column 5 is p value.

	1-Df	2-Sum Sq	3-Mean Sq	4-F value	5-Pr(>F)
strains	28	14.09402	0.503358	62.5602	7.91E-34
Residuals	58	0.466667	0.008046		

Significance codes: '***' 0.001

Table 3.8. Multi-comparison of chloramphenicol resistance of 30 *A. tumefaciens* strains. Column 1-strain name. Column 2 is the estimated mean of the inhibition zone of each train based on the model. Column 3 is standard error. Column 4 is degrees of freedom. Columns 5 and 6 are the lower confidence level and upper confidence level respectively. Column 7 represents the statistical group in terms of significance. Confidence level = 0.95. p-value adjustment: Tukey method for comparing a family of 30 estimates. Significance level = 0.05.

1-strains	2-emmean	3-SE	4-df	5-lower.CL	6-upper.CL	7-.group
Kin003	0.7	0.05	58	0.6	0.8	a
Tul003	0.73	0.05	58	0.63	0.84	ab
Tul004	0.77	0.05	58	0.66	0.87	abc
Teh001	0.8	0.05	58	0.7	0.9	abcd
Sta003	0.83	0.05	58	0.73	0.94	abcde
SJ003	0.87	0.05	58	0.76	0.97	abcdef
But002	0.93	0.05	58	0.83	1.04	abcdefg
Tul001	0.93	0.05	58	0.83	1.04	abcdefg
Teh002	0.93	0.05	58	0.83	1.04	abcdefg
Sut002	0.93	0.05	58	0.83	1.04	abcdefg
Sta005	0.97	0.05	58	0.86	1.07	abcdefg
CC001	0.97	0.05	58	0.86	1.07	abcdefg
Sut001	0.97	0.05	58	0.86	1.07	abcdefg
Gle002	1	0.05	58	0.9	1.1	bcdefg
Sta002	1.03	0.05	58	0.93	1.14	cdefg
Yol001	1.03	0.05	58	0.93	1.14	cdefg
Yol002	1.03	0.05	58	0.93	1.14	cdefg
Gle001	1.03	0.05	58	0.93	1.14	cdefg
SJ001	1.07	0.05	58	0.96	1.17	defg
Sta001	1.07	0.05	58	0.96	1.17	defg
CL001	1.07	0.05	58	0.96	1.17	defg
SJ002	1.07	0.05	58	0.96	1.17	defg
Yub001	1.1	0.05	58	1	1.2	efg
Kin001	1.13	0.05	58	1.03	1.24	fg
Yub002	1.17	0.05	58	1.06	1.27	g
C58	1.5	0.05	58	1.4	1.6	h
Tul002	1.9	0.05	58	1.8	2	i
Sta004	2.2	0.05	58	2.1	2.3	j
But001	2.47	0.05	58	2.36	2.57	j

Confidence level used: 0.95
P value adjustment: Tukey method for comparing a family of 30 estimates
Significance level used: alpha = 0.05

Table 3.9. Analysis of variance of ciprofloxacin resistance of 30 *A. tumefaciens* strains. Column 1 is degrees of freedom (Df). Column 2 is Sum of squares (Sum-Sq). Column 3 is Mean Squares (Mean-Sq). Column 4 is F ratio. Column 5 is p value.

	1-Df	2-Sum Sq	3-Mean Sq	4-F value	5-Pr(>F)
strains	29	14.1891	0.489279	34.73898	9.57E-28
Residuals	60	0.845067	0.014084		

Signif. codes: '***' 0.001

Table 3.10. Multi-comparison of ciprofloxacin resistance of 30 *A. tumefaciens* strains. Column 1 is the strain name. Column 2 is the estimated mean of the inhibition zone of each train based on the model. Column 3 is standard error. Column 4 is degrees of freedom. Columns 5 and 6 are the lower confidence level and upper confidence level respectively. Column 7 represents the statistical group in terms of significance. Confidence level = 0.95. p-value adjustment: Tukey method for comparing a family of 30 estimates. Significance level = 0.05.

1-strains	2-emmean	3-SE	4-df	5-lower.CL	6-upper.CL	7-.group
Kin002	1.9	0.07	60	1.76	2.04	a
Tul002	2.57	0.07	60	2.43	2.7	b
Kin003	2.63	0.07	60	2.5	2.77	bc
Gle001	2.67	0.07	60	2.53	2.8	bcd
SJ003	2.7	0.07	60	2.57	2.84	bcde
Sta002	2.83	0.07	60	2.7	2.97	bcdef
But001	2.9	0.07	60	2.76	3.04	bcdefg
Yub002	3	0.07	60	2.86	3.14	cdefg
CL001	3.03	0.07	60	2.9	3.17	defgh
Sta004	3.03	0.07	60	2.9	3.17	defgh
But002	3.03	0.07	60	2.9	3.17	defgh
SJ001	3.03	0.07	60	2.9	3.17	defgh
Sut002	3.07	0.07	60	2.93	3.2	efghi
Teh001	3.1	0.07	60	2.96	3.24	fghi
CC001	3.1	0.07	60	2.96	3.24	fghi
Sta003	3.1	0.07	60	2.96	3.24	fghi
Teh002	3.13	0.07	60	3	3.27	fghij
Sta001	3.2	0.07	60	3.06	3.34	fghijk
Sut001	3.23	0.07	60	3.1	3.37	ghijkl
Kin001	3.4	0.07	60	3.26	3.54	hijklm
Sta005	3.43	0.07	60	3.3	3.57	ijklm
Tul001	3.43	0.07	60	3.3	3.57	ijklm
SJ002	3.43	0.07	60	3.3	3.57	ijklm
Tul003	3.5	0.07	60	3.36	3.64	jklmna
Yol002	3.5	0.07	60	3.36	3.64	jklmna
Tul004	3.53	0.07	60	3.4	3.67	klmna
Gle002	3.6	0.07	60	3.46	3.74	lmna
Yol001	3.63	0.07	60	3.5	3.77	mna
Yub001	3.67	0.07	60	3.53	3.8	mna
C58	3.83	0.07	60	3.7	3.97	na

Confidence level used: 0.95
P value adjustment: Tukey method for comparing a family of 30 estimates
Significance level used: alpha = 0.05

Table 3.11. Analysis of variance of erythromycin resistance of 30 *A. tumefaciens* strains. Column 1 is degrees of freedom(Df). Column 2 is Sum of squares (Sum-Sq). Column 3 is Mean Squares (Mean-Sq). Column 4 is F ratio. Column 5 is p value.

	1-Df	2-Sum Sq	3-Mean Sq	4-F value	5-Pr(>F)
strains	29	4.777889	0.164755	18.30609	2.56E-20
Residuals	60	0.54	0.009		

Signif. codes: '***' 0.001

Table 3.12. Multi-comparison of erythromycin resistance of 30 *A. tumefaciens* strains. Column 1 is the strain name. Column 2 is the estimated mean of the inhibition zone of each train based on the model. Column 3 is standard error. Column 4 is degrees of freedom. Columns 5 and 6 are the lower confidence level and upper confidence level, respectively. Column 7 represents the statistical group in terms of significance. Confidence level = 0.95. p-value adjustment: Tukey method for comparing a family of 30 estimates. Significance level = 0.05.

1-strains	2-emmean	3-SE	4-df	5-lower.CL	6-upper.CL	7.group
But002	0.7	0.05	60	0.59	0.81	a
Kin002	0.77	0.05	60	0.66	0.88	ab
Sta001	0.87	0.05	60	0.76	0.98	abc
Sta005	0.87	0.05	60	0.76	0.98	abc
Sut001	0.9	0.05	60	0.79	1.01	abcd
Gle001	0.93	0.05	60	0.82	1.04	abcde
Kin001	0.93	0.05	60	0.82	1.04	abcde
Tul003	0.97	0.05	60	0.86	1.08	abcde
SJ002	0.97	0.05	60	0.86	1.08	abcde
Gle002	0.97	0.05	60	0.86	1.08	abcde
Yol001	1	0.05	60	0.89	1.11	abcde
Sta003	1	0.05	60	0.89	1.11	abcde
But001	1.03	0.05	60	0.92	1.14	bcde
Teh001	1.07	0.05	60	0.96	1.18	bcdef
Tul004	1.07	0.05	60	0.96	1.18	bcdef
Yol002	1.07	0.05	60	0.96	1.18	bcdef
Yub002	1.07	0.05	60	0.96	1.18	bcdef
SJ001	1.07	0.05	60	0.96	1.18	bcdef
Tul001	1.1	0.05	60	0.99	1.21	cdef
Kin003	1.1	0.05	60	0.99	1.21	cdef
Sta002	1.1	0.05	60	0.99	1.21	cdef
CL001	1.13	0.05	60	1.02	1.24	cdefg
Teh002	1.13	0.05	60	1.02	1.24	cdefg
Sta004	1.17	0.05	60	1.06	1.28	cdefg
SJ003	1.2	0.05	60	1.09	1.31	defg
CC001	1.23	0.05	60	1.12	1.34	efg
Sut002	1.23	0.05	60	1.12	1.34	efg
Tul002	1.37	0.05	60	1.26	1.48	fg
C58	1.43	0.05	60	1.32	1.54	g
Yub001	2	0.05	60	1.89	2.11	h

Confidence level used: 0.95
P value adjustment: Tukey method for comparing a family of 30 estimates
Significance level used: alpha = 0.05

Table 3.13. Analysis of variance of rifampin resistance of 30 *A. tumefaciens* strains. Column 1 is degrees of freedom(Df). Column 2 is Sum of squares (Sum-Sq). Column 3 is Mean Squares (Mean-Sq). Column 4 is F ratio. Column 5 is p value.

	1-Df	2-Sum Sq	3-Mean Sq	4-F value	5-Pr(>F)
strains	28	1.226839	0.043816	10.2693	8.55E-14
Residuals	58	0.247467	0.004267		

Signif. codes: '***' 0.001

Table 3.14. Multi-comparison of rifampin resistance of 30 *A. tumefaciens* strains. Column 1 is the strain name. Column 2 is the estimated mean of the inhibition zone of each train based on the model. Column 3 is standard error. Column 4 is degrees of freedom. Columns 5 and 6 are the lower confidence level and upper confidence level respectively. Column 7 represents the statistical group in terms of significance. Confidence level = 0.95. p-value adjustment: Tukey method for comparing a family of 30 estimates. Significance level = 0.05.

1-strains	2-emmean	3-SE	4-df	5-lower.CL	6-upper.CL	7-.group
Sta004	1.13	0.04	58	1.06	1.21	a
Sta005	1.17	0.04	58	1.09	1.24	ab
Sut001	1.17	0.04	58	1.09	1.24	ab
Gle001	1.17	0.04	58	1.09	1.24	ab
Sta001	1.2	0.04	58	1.12	1.28	abc
But002	1.23	0.04	58	1.16	1.31	abcd
Kin002	1.23	0.04	58	1.16	1.31	abcd
Teh001	1.23	0.04	58	1.16	1.31	abcd
Tul003	1.23	0.04	58	1.16	1.31	abcd
Tul004	1.23	0.04	58	1.16	1.31	abcd
SJ001	1.25	0.04	58	1.17	1.33	abcd
But001	1.27	0.04	58	1.19	1.34	abcd
CL001	1.27	0.04	58	1.19	1.34	abcd
SJ003	1.27	0.04	58	1.19	1.34	abcd
Yol001	1.27	0.04	58	1.19	1.34	abcd
Gle002	1.3	0.04	58	1.22	1.38	abcde
Tul001	1.3	0.04	58	1.22	1.38	abcde
SJ002	1.32	0.04	58	1.24	1.39	abcde
CC001	1.33	0.04	58	1.26	1.41	abcdef
Teh002	1.33	0.04	58	1.26	1.41	abcdef
Yol002	1.33	0.04	58	1.26	1.41	abcdef
Sut002	1.37	0.04	58	1.29	1.44	bcdef
Yub001	1.4	0.04	58	1.32	1.48	cdefg
Sta002	1.43	0.04	58	1.36	1.51	defg
Kin001	1.43	0.04	58	1.36	1.51	defg
Yub002	1.5	0.04	58	1.42	1.58	efg
Tul002	1.53	0.04	58	1.46	1.61	fg
Kin003	1.53	0.04	58	1.46	1.61	fg
C58	1.6	0.04	58	1.52	1.68	g

Confidence level used: 0.95
P value adjustment: tukey method for comparing a family of 30 estimates
Significance level used: alpha = 0.05

Table 3.15. Analysis of variance of tetracycline resistance of 30 *A. tumefaciens* strains. Column 1 is degrees of freedom(Df). Column 2 is Sum of squares (Sum-Sq). Column 3 is Mean Squares (Mean-Sq). Column 4 is F ratio. Column 5 is p value.

	1-Df	2-Sum Sq	3-Mean Sq	4-F value	5-Pr(>F)
strains	29	14.71303	0.507346	39.44464	2.83E-29
Residuals	60	0.771733	0.012862		

Signif. codes: '***' 0.001

Table 3.16. Multi-comparison of tetracycline resistance of 30 *A. tumefaciens* strains. Column 1 is the strain name. Column 2 is the estimated mean of the inhibition zone of each train based on the model. Column 3 is standard error. Column 4 is degrees of freedom. Columns 5 and 6 are the lower confidence level and upper confidence level respectively. Column 7 represents the statistical group in terms of significance. Confidence level = 0.95. p-value adjustment: Tukey method for comparing a family of 30 estimates. Significance level = 0.05.

1-strains	2-emmean	3-SE	4-df	5-lower.CL	6-upper.CL	7-.group
Sta001	2.23	0.07	60	2.1	2.36	a
But002	2.27	0.07	60	2.14	2.4	ab
Sut001	2.27	0.07	60	2.14	2.4	ab
Gle001	2.37	0.07	60	2.24	2.5	abc
Tul004	2.43	0.07	60	2.3	2.56	abc
SJ001	2.45	0.07	60	2.32	2.58	abc
Sta002	2.47	0.07	60	2.34	2.6	abc
SJ002	2.5	0.07	60	2.37	2.63	abcd
Sut002	2.5	0.07	60	2.37	2.63	abcd
Tul001	2.5	0.07	60	2.37	2.63	abcd
Teh001	2.6	0.07	60	2.47	2.73	bcde
Teh002	2.63	0.07	60	2.5	2.76	cdef
Yol001	2.7	0.07	60	2.57	2.83	cdefg
Gle002	2.7	0.07	60	2.57	2.83	cdefg
Sta003	2.83	0.07	60	2.7	2.96	defgh
Yol002	2.87	0.07	60	2.74	3	efgh
SJ003	2.93	0.07	60	2.8	3.06	efghi
Kin003	2.97	0.07	60	2.84	3.1	fghi
Sta005	3.03	0.07	60	2.9	3.16	ghik
CC001	3.03	0.07	60	2.9	3.16	ghik
Tul002	3.03	0.07	60	2.9	3.16	ghik
Kin001	3.07	0.07	60	2.94	3.2	hikl
CL001	3.1	0.07	60	2.97	3.23	hikl
Tul003	3.1	0.07	60	2.97	3.23	hikl
Yub002	3.27	0.07	60	3.14	3.4	ikl
Sta004	3.33	0.07	60	3.2	3.46	klm
Kin002	3.37	0.07	60	3.24	3.5	klmn
But001	3.4	0.07	60	3.27	3.53	lmn
C58	3.63	0.07	60	3.5	3.76	mn
Yub001	3.7	0.07	60	3.57	3.83	n

Confidence level used: 0.95
P value adjustment: tukey method for comparing a family of 30 estimates
Significance level used: alpha = 0.05

Table 3.17. Analysis of variance of K84 sensitivity of 15 *A. tumefaciens* strains. Column 1 is degrees of freedom (Df). Column 2 is Sum of squares (Sum-Sq). Column 3 is Mean Squares (Mean-Sq). Column 4 is F ratio. Column 5 is p value.

	1-Df	2-Sum Sq	3-Mean Sq	4-F value	5-Pr(>F)
strains	14	18.9287	1.35205	108.1255	1.03E-21
Residuals	30	0.375133	0.012504		

Signif. codes: '***' 0.001

Table 3.18. Multi-comparison of agrocin-84 sensitivity of 15 *A. tumefaciens* strains. Column 1 is the strain name. Column 2 is the estimated mean of the inhibition zone of each train based on the model. Column 3 is standard error. Column 4 is degrees of freedom. Columns 5 and 6 are the lower confidence level and upper confidence level respectively. Column 7 represents the statistical group in terms of significance. Confidence level = 0.95. p-value adjustment: Tukey method for comparing a family of 15 estimates. Significance level = 0.05.

1-strains	2-emmean	3-SE	4-df	5-lower.CL	6-upper.CL	7-.group
Tul003	2.87	0.06	30	2.73	3	a
Yub001	4.51	0.06	30	4.38	4.64	b
Sta005	4.52	0.06	30	4.39	4.65	b
CC001	4.82	0.06	30	4.68	4.95	bc
Sta003	5.11	0.06	30	4.98	5.24	cd
Tul001	5.13	0.06	30	5	5.27	cde
Sta001	5.28	0.06	30	5.15	5.41	def
Kin003	5.29	0.06	30	5.16	5.43	def
Yol001	5.31	0.06	30	5.18	5.45	def
Yol002	5.31	0.06	30	5.18	5.45	def
But001	5.32	0.06	30	5.18	5.45	def
C58	5.34	0.06	30	5.21	5.48	def
Sta004	5.41	0.06	30	5.27	5.54	def
SJ002	5.45	0.06	30	5.32	5.59	ef
SJ001	5.53	0.06	30	5.39	5.66	f

Confidence level used: 0.95
P value adjustment: Tukey method for comparing a family of 15 estimates
Significance level used: alpha = 0.05

Table 3.19. Analysis of variance of motility of 30 *A. tumefaciens* strains. Column 1 is degrees of freedom(Df). Column 2 is Sum of squares (Sum-Sq). Column 3 is Mean Squares (Mean-Sq). Column 4 is F ratio. Column 5 is p value.

	1-Df	2-Sum Sq	3-Mean Sq	4-F value	5-Pr(>F)
strains	29	7.1722	0.247317	48.30415	9.63E-32
Residuals	60	0.3072	0.00512		

Signif. codes: '***' 0.001

Table 3.20. Multi-comparison of motility of 30 *A. tumefaciens* strains. Column 1 is the strain name. Column 2 is the estimated mean of the inhibition zone of each train based on the model. Column 3 is standard error. Column 4 is degrees of freedom. Columns 5 and 6 are the lower confidence level and upper confidence level respectively. Column 7 represents the statistical group in terms of significance. Confidence level = 0.95. p-value adjustment: Tukey method for comparing a family of 30 estimates. Significance level = 0.05.

1-strains	2-emmean	3-SE	4-df	5-lower.CL	6-upper.CL	7-.group
CL001	3.18	0.04	60	3.09	3.26	a
Kin002	3.26	0.04	60	3.18	3.35	ab
Yub002	3.27	0.04	60	3.19	3.36	abc
But001	3.33	0.04	60	3.25	3.41	abcd
Tul003	3.39	0.04	60	3.31	3.47	abcde
Tul002	3.44	0.04	60	3.36	3.53	bcdef
Teh002	3.46	0.04	60	3.38	3.55	bcdefg
Gle002	3.49	0.04	60	3.41	3.57	bcdefgh
But002	3.5	0.04	60	3.41	3.58	cdefgh
Teh001	3.5	0.04	60	3.42	3.58	cdefgh
Kin001	3.5	0.04	60	3.42	3.59	defgh
Sut001	3.51	0.04	60	3.43	3.59	defgh
Tul001	3.51	0.04	60	3.43	3.59	defgh
Kin003	3.59	0.04	60	3.5	3.67	efghi
C58	3.59	0.04	60	3.51	3.68	efghi
Gle001	3.62	0.04	60	3.54	3.7	fghik
Sta005	3.63	0.04	60	3.54	3.71	fghik
CC001	3.69	0.04	60	3.6	3.77	ghikl
Sta003	3.71	0.04	60	3.62	3.79	hikl
Sut002	3.71	0.04	60	3.63	3.79	hikl
Tul004	3.75	0.04	60	3.66	3.83	ikl
Sta001	3.75	0.04	60	3.67	3.84	ikl
Sta002	3.76	0.04	60	3.68	3.85	ikl
SJ002	3.81	0.04	60	3.72	3.89	ikl
Yol001	3.85	0.04	60	3.76	3.93	klm
SJ001	3.88	0.04	60	3.79	3.96	lm
Yol002	4.07	0.04	60	3.98	4.15	mn
Sta004	4.11	0.04	60	4.03	4.19	n
SJ003	4.19	0.04	60	4.11	4.28	n
Yub001	4.46	0.04	60	4.37	4.54	o

Confidence level used: 0.95
P value adjustment: Tukey method for comparing a family of 30 estimates
Significance level used: alpha = 0.05

Table 3.21. Pearson correlation matrix of ten phenotypic traits (variables). The same varibales show in both rows and columns. They are motility (mot_mean), carbenicillin resistance (cb_mean), chloramphenicol resistance (chl_mean), ciprofloxacin resistance(cip_mean), erythromycin resistance (ery_mean), rifampin resistance (rif_mean), tetracycline resistance (tet_mean), K84 sensitivity (k84_mean), growth rate in TSB medium(tsb_mean), and growth rate in AB medium (ab_mean). The diagonal coefficients shows that each variable perfectly correlates with itself. No obvious correlation was observed between either of the two variables. The matrix is symmetrical, with the same correlation is shown above the main diagonal being a mirror image of those below the main diagonal.

Dtumor_means	mot_mean	cb_mean	chl_mean	cip_mean	ery_mean	rif_mean	tet_mean	k84_mean	tsb_mean	ab_mean
1	0.23	0	0.11	0.62	0.06	-0.17	-0.02	0.63	0.02	-0.11
0.23	1	-0.38	0.09	0.28	0.5	-0.07	0.03	0.4	0.24	-0.19
0	-0.38	1	0.05	0.16	-0.5	-0.19	-0.15	0.09	-0.2	0.38
0.11	0.09	0.05	1	0.13	0.28	0.14	0.31	0.28	0.27	-0.17
0.62	0.28	0.16	0.13	1	0.25	0.04	-0.01	0.36	0.4	0.11
0.06	0.5	-0.5	0.28	0.25	1	0.28	0.52	0.18	0.3	0.05
-0.17	-0.07	-0.19	0.14	0.04	0.28	1	0.18	-0.17	0.32	0.23
-0.02	0.03	-0.15	0.31	-0.01	0.52	0.18	1	0.23	0.2	0.08
0.63	0.4	0.09	0.28	0.36	0.18	-0.17	0.23	1	-0.03	-0.02
0.02	0.24	-0.2	0.27	0.4	0.3	0.32	0.2	-0.03	1	-0.14
-0.11	-0.19	0.38	-0.17	0.11	0.05	0.23	0.08	-0.02	-0.14	1

References

1. Cleene, M. D. The Susceptibility of Monocotyledons to *Agrobacterium tumefaciens*. *Journal of Phytopathology* **113**, 81–89 (1985).
2. Hooykaas, Paul. J. J. & Beijersbergen, A. G. M. The Virulence System of *Agrobacterium Tumefaciens*. *Annual Review of Phytopathology* **32**, 157–181 (1994).
3. Shaw, C. H. *et al.* virA and virG are the Ti-plasmid functions required for chemotaxis of *Agrobacterium tumefaciens* towards acetosyringone. *Molecular Microbiology* **2**, 413–417 (1988).
4. Harshey, R. M. Bacterial Motility on a Surface: Many Ways to a Common Goal. *Annual Review of Microbiology* **57**, 249–273 (2003).
5. Abu-Ashour, J., Joy, D. M., Lee, H., Whiteley, H. R. & Zelin, S. Transport of microorganisms through soil. *Water Air Soil Pollut* **75**, 141–158 (1994).
6. Tans-Kersten, J., Huang, H. & Allen, C. Ralstonia solanacearum Needs Motility for Invasive Virulence on Tomato. *Journal of Bacteriology* **183**, 3597–3605 (2001).
7. Merritt, P. M., Danhorn, T. & Fuqua, C. Motility and Chemotaxis in *Agrobacterium tumefaciens* Surface Attachment and Biofilm Formation. *Journal of Bacteriology* **189**, 8005–8014 (2007).
8. Josenhans, C. & Suerbaum, S. The role of motility as a virulence factor in bacteria. *International Journal of Medical Microbiology* **291**, 605–614 (2002).
9. Pradhanang, P. M., Momol, M. T., Olson, S. M. & Jones, J. B. Effects of Plant Essential Oils on *Ralstonia solanacearum* Population Density and Bacterial Wilt Incidence in Tomato. *Plant Disease* **87**, 423–427 (2003).
10. Ram, Y. *et al.* Predicting microbial growth in a mixed culture from growth curve data. *PNAS* **116**, 14698–14707 (2019).
11. ScienceDirect Full Text PDF.

12. The Disease Triangle: A plant pathological paradigm revisited. https://www.apsnet.org/edcenter/foreducators/TeachingNotes/Pages/DiseaseTriangle.aspx.
13. Allen, R. J. & Waclaw, B. Bacterial growth: a statistical physicist's guide. *Rep Prog Phys* **82**, 016601 (2019).
14. Stewart, P. S. Biofilm accumulation model that predicts antibiotic resistance of *Pseudomonas aeruginosa* biofilms. *Antimicrobial Agents and Chemotherapy* **38**, 1052–1058 (1994).
15. Bauer, M. A., Kainz, K., Carmona-Gutierrez, D. & Madeo, F. Microbial wars: Competition in ecological niches and within the microbiome. *Microb Cell* **5**, 215–219.
16. Zaman, S. B. *et al.* A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus* **9**,.
17. Sundin, G. W. & Wang, N. Antibiotic Resistance in Plant-Pathogenic Bacteria. *Annual Review of Phytopathology* **56**, 161–180 (2018).
18. Penyalver, R., Vicedo, B. & López, M. M. Use of the Genetically Engineered *Agrobacterium* Strain K1026 for Biological Control of Crown Gall. *European Journal of Plant Pathology* **106**, 801–810 (2000).
19. Ryder, M. H. & Jones, D. A. Biological Control of Crown Gall Using Using *Agrobacterium* Strains K84 and K1026. *Functional Plant Biol.* **18**, 571–579 (1991).
20. Vladimirov, I. A., Matveeva, T. V. & Lutova, L. A. Opine biosynthesis and catabolism genes of *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*. *Russ J Genet* **51**, 121–129 (2015).
21. Yakabe, L. E. *et al.* Novel PCR primers for detection of genetically diverse virulent *Agrobacterium tumefaciens* biovar 1 strains. *J Gen Plant Pathol* **78**, 121–126 (2012).
22. Yakabe, L. E., Parker, S. R. & Kluepfel, D. A. Incidence of *Agrobacterium tumefaciens* Biovar 1 in and on 'Paradox' (*Juglans hindsii* × *Juglans regia*) Walnut Seed Collected from Commercial Nurseries. *Plant Disease* **98**, 766–770 (2014).

23. Sundin, G. W. & Wang, N. Antibiotic Resistance in Plant-Pathogenic Bacteria. *Annual Review of Phytopathology* **56**, 161–180 (2018).
24. Stonier, T. *AGROBACTERIUM TUMEFACIENS* CONN II. *J Bacteriol* **79**, 889–898 (1960).
25. McClean, A. E. & Kluepfel, D. A. Genetic Loci Involved in Rubrifacine Production in the Walnut Pathogen *Brenneria rubrifaciens*. *Phytopathology®* **99**, 145–151 (2009).
26. Sprouffske, K. & Wagner, A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics* **17**, 172 (2016).
27. Grolemund, G. & Wickham, H. Dates and Times Made Easy with lubridate. *Journal of Statistical Software* **40**, 1–25 (2011).
28. Wickham, H., François, R., Henry, L., Müller, K. & RStudio. *dplyr: A Grammar of Data Manipulation*. (2020).
29. Henry, L., Wickham, H. & RStudio. *tidyselect: Select from a Set of Strings*. (2020).
30. R: The R Project for Statistical Computing. https://www.r-project.org/.
31. lme4.pdf.
32. Wickham, H. ggplot2. *WIREs Computational Statistics* **3**, 180–185 (2011).
33. Lenth, R. V. *et al. emmeans: Estimated Marginal Means, aka Least-Squares Means*. (2020).
34. rcompanion.pdf.
35. car.pdf.
36. Hothorn, T. *et al. multcomp: Simultaneous Inference in General Parametric Models*. (2020).
37. Kado, C. I. Historical account on gaining insights on the mechanism of crown gall tumorigenesis induced by *Agrobacterium tumefaciens*. *Front Microbiol* **5**, (2014).
38. Möller, P. *et al.* Profound Impact of Hfq on Nutrient Acquisition, Metabolism and Motility in the Plant Pathogen *Agrobacterium tumefaciens*. *PLOS ONE* **9**, e110427 (2014).

39. MOLLER, W., WJ, M., MN, S. & SV, T. THE SCENARIO OF FIRE BLIGHT AND STREPTOMYCIN RESISTANCE. *THE SCENARIO OF FIRE BLIGHT AND STREPTOMYCIN RESISTANCE* (1981).
40. Adaskaveg, L. M. | J. E. New Crop Protection Tool in Walnut Blight Fight. *Growing Produce* https://www.growingproduce.com/nuts/new-crop-protection-tool-walnut-blight-fight/ (2018).
41. Benveniste, R. & Davies, J. Aminoglycoside Antibiotic-Inactivating Enzymes in Actinomycetes Similar to Those Present in Clinical Isolates of Antibiotic-Resistant Bacteria. *Proc Natl Acad Sci U S A* **70**, 2276–2280 (1973).
42. Chopra, I. & Roberts, M. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol Mol Biol Rev* **65**, 232-260 ; second page, table of contents (2001).
43. Grossman, T. H. Tetracycline Antibiotics and Resistance. *Cold Spring Harb Perspect Med* **6**, a025387 (2016).
44. Vicedo, B., Peñalver, R., Asins, M. J. & López, M. M. Biological Control of *Agrobacterium tumefaciens*, Colonization, and pAgK84 Transfer with *Agrobacterium radiobacter* K84 and the Tra- Mutant Strain K1026. *Appl. Environ. Microbiol.* **59**, 309–315 (1993).
45. López, M. M., Gorris, M. T., Salcedo, C. I., Montojo, A. M. & Miró, M. Evidence of Biological Control of *Agrobacterium tumefaciens* Strains Sensitive and Resistant to Agrocin 84 by Different Agrobacterium radiobacter Strains on Stone Fruit Trees. *Appl Environ Microbiol* **55**, 741–746 (1989).
46. Basra, P. *et al.* Fitness Tradeoffs of Antibiotic Resistance in Extraintestinal Pathogenic Escherichia coli. *Genome Biol Evol* **10**, 667–679 (2018).
47. Bélanger, C., Canfield, M. L., Moore, L. W. & Dion, P. Genetic analysis of nonpathogenic *Agrobacterium tumefaciens* mutants arising in crown gall tumors. *Journal of bacteriology* **177**, 3752–3757 (1995).

Chapter 4

A comparative genomics analysis reveals a potential genetic basis for key survival phenotypes in *Agrobacterium tumefaciens*

Abstract

Fifteen ANI groups were identified after examining the genomic DNA sequence data from 311 *Agrobacterium tumefaciens* strains. A subset of these strains, consisting of 28 *Agrobacterium* strains (25 virulent and 3 avirulent) isolated from crown gall tissue and soil collected from the top ten walnut producing counties in California were placed into eight ANI groups. A comparative genomic analysis of these 28 *A. tumefaciens* strains, plus A. tumefaciens straisn C58 and CL001, revealed synteny of genetic elements involved in pathogenesis. These elements include highly conserved T4SS, T-DNA, and the *vir*A/*vir*G two component regulatory system. Based on their opine biosynthesis genes, this group of strains was further subdivided into agropine, succinamopine, and nopaline-type strains. The right-border sequences of 27 strains (25 virulent plus C58 and CL001) were paired with a specific opine-type. The agropine and succinamopine opine types were correlated to the size of tumors induced on *Datura stramonium*, but not on walnut. The different hosts (*D. stramonium* and walnut) also had a significant influence on the tumors induced by a given *A. tumefaciens* opine genotype. These data indicate that the opine type may influence a strain's host preference. I also examined the T6SS of these 30 *A. tumefaciens* strains. Three virulent strains did not contain a T6SS. The remaining strains, i.e., both virulent and avirulent, contained a classical T6SS in which the *imp* operon was conserved, and the *hcp* operons were found to be variable. I was able to classify the *hcp* operons into eight unique *hcp*-clusters. Variability in the *hcp* operon may contribute to the ability of *A. tumefaciens* to adapt to a wide range of environmental conditions.

Keywords: Agrobacterium tumefaciens, virulence, T6SS, opine-types, comparative genomics

Introduction

Agrobacterium tumefaciens is a soil-borne bacterial pathogen that causes crown gall (CG) disease on most dicotyledonous plants[1]. Crown Gall frequently plagues United States agriculture where it can reduce yields of a wide variety of woody perennial plants, e.g. walnut, almond, grapes, rose, and

numerous ornamentals [2]. Due to the challenging nature of CG management, many control strategies have been explored. Significantly, the biocontrol strain, *Agrobacterium rhizogenes* K84 [3], first reported in 1972 by New and Kerr [4], showed great promise for controlling CG. However, many virulent *A. tumefaciens* strains have been identified resistant to strain K84. Currently, there is no commercially available biological control agent or bactericide that is effective or reliable for pre- or post-plant control of CG on woody perennial crops.

Understanding the diversity of virulent *A. tumefaciens* strains may facilitate development of new and novel CG disease management strategies. However, our current understanding of *A. tumefaciens* revolves around our knowledge of a limited number of strains, such as C58, Ach5, and A6. [5]. The phylogenetic analyses of 311 *Agrobacterium* strains based on average nucleotide identity (ANI) and core gene analyses identified 15 ANI groups and at least ten genomic species in *A. tumefaciens.* However, these conclusions are based on a small number of strains and should be view with caution (Chapter 2). Consequently, it is critical that this high level of genetic diversity be understood and taken into consideration when addressing such issues as CG management strategies, breeding for CG resistance, and development of broad-spectrum biological control agents.

Infecting plant hosts to induce host production of nutrients and energy is one strategy adopted by *A. tumefaciens* to survive in an adverse environment [6]. Other strategies to promote survival and competition include secreting antimicrobial compounds, such as antibiotics, to inhibit other soil-borne microbes [7,8,9]. Antibiotics also have been exploited by humans to manage CG and other harmful pathogens but the overuse of antibiotics often leads to the generation of antibiotic-resistance microbial pathogens [10]. Though the application of antibiotics in agriculture is regulated to limit selection of resistant strains, microbes can acquire an antibiotic-resistant phenotype through a variety of molecular mechanisms, such as horizontal gene transfer (HGT) [11]. Currently, there are no data available on how *A. tumefaciens*

responds to exposure to various antibiotics in terms of the spread or degree to which antibiotic resistant strains develop in the population.

Another defense strategy used by Gram-negative bacteria is to deploy toxins or effectors into microbe's cells via a type VI secretion system (T6SS). (Note; *A. tumefaciens* does use a T3SS for this)[12]. T6SS is a widely conserved multicomponent nanomachine that is structurally related to contractile phage tail-like structures [13,14]. In *A. tumefaciens* strain C58, the T6SS-encoding locus comprises two divergently transcribed operons. One is the *imp* operon, which consists of 14 genes (*atu*4343 to *atu*4330) [15], eleven of which are considered to be core T6SS genes in Proteobacteria [16]. The other is the *hcp* operon encoding 9 genes (*atu*4344 to *atu*4352) [15]. The type VI secretion amidase effector (*tae, atu4346*) - type VI secretion amidase immunity (*tai*, *atu4347*) pairs and type VI DNase effector (*tde*, *atu4350*) – type VI DNase immunity (*tdi, atu4351*) pairs are in the *hcp* operon. The *tde2* (*atu3640*) – *tdi2* (*atu3639*) pairs are in the VgrG2 module. Tae may be required for the delivery of Tde1 and Tde2, which are toxic to other microbes [17].

A study evaluating the role of the T6SS in 11 *A. tumefaciens* strains representing four genomic species reported that the *imp* loci are conserved, whereas genes encoding EI pairs downstream of *vgr*G in the *hcp* loci are variable among members of these four genomospecies [18]. In summary, the authors' findings suggest a role of the T6SS expression in microbial antagonism (i.e., secreting toxic effectors), which impacts microbial population and community dynamics. However, T6SS-dependent bacterial antagonism was not predictable based on the presence of T6SS genetics/machinery [17]. This suggests that other genetic and ecological factors may have significant impacts on T6SS activity [17]. Again, this provides further justification to consider a global genomic approach when examining phenotypic properties of exhibited members of the *A. tumefaciens* complex [18].

I previously reported that 29 (28 from CA walnut orchards + one from Chile walnut orchard) *A. tumefaciens* strains isolated from walnut orchards were genetically very diverse based on whole genome

phylogeny analysis (Chapter 2). Thus, these strains represented ten genomic species (also ten ANI groups). To further investigate this genetic diversity and relate it to the observed phenotypic diversity, I performed in silico comparative genomic analyses to reveal the potential genetic basis underlying the various strategies related to *A. tumefaciens* fitness in the adverse soil environment. To accomplish this objective, I generated and used high-quality complete genomes from 29 (28 from CA walnut orchards + one from Chile walnut orchard) *A. tumefaciens* strains (26 virulent and 3 non-virulent), which I isolated from CG tissues and soil samples from California walnut orchards and nurseries.

Here, I examined the genetic basis of the observed phenotypic diversity described in Chapter 3. I observed a statistically significant difference in terms of the virulence, antibiotic resistance, growth rate and K84 sensitivity among the 29 *A. tumefaciens* strains (Chapter 3). Using comparative genomics, I identified three types of Ti plasmids as well as unique transfer DNA (T-DNA) right border sequences. The genetic basis of motility was not analyzed in this chapter. The *acc* operon in various Ti plasmids also exhibited significant polymorphism, which may explain why some virulent *A. tumefaciens* strains are resistant to biocontrol strain K84. I also observed great genetic diversity in the T6SS and their predicted T6SS effectors. Interestingly, three virulent strains were found to lack a T6SS, which is consistent with previous findings that the T6SS is not required for pathogenicity [18]. Collectively, these data provide a comprehensive insight into genetic loci that are associated with key phenotypes in the *A. tumefaciens* population in walnut orchards in California. This work may be useful for walnut rootstock breeders when breeding for CG resistance. It also provides new insights for plant pathologists and growers as they develop new sustainable methods for crown gall disease management.

Methods

Twenty-nine *A. tumefaciens* strains were isolated from CG tissues or soil next to CG tissues, single colony purified, and preserved at -80 °C. The phenotypic data were obtained as described in Chapter 3. The genomic sequencing, assembly, and annotation of the 29 *A. tumefaciens* strains from walnut is

described in Chapter 2. The *A. tumefaciens* type strain C58 was included as a reference. In total, 30 *A. tumefaciens* complete genomes were included in this study (29 walnut strains plus *A. tumefaciens* C58; 27 virulent; 3 avirulent).

Phylogenomic tree construction

To generate the phylogeny of the 30 *A. tumefaciens* strains used in Fig. 4.18, I performed a pangenome analysis using PIRATEv1.0.4 tool [19] to obtain the core genes alignment. Using FastTreev2.1.10 [20], I inferred an approximately-maximum-likelihood phylogenetic tree under the Generalized Time-Reversible model computing local support values with the Shimodaira-Hasegawa test at SH-like 1000. The tree was rooted at strain Yub001 (Chapter 2).

Because the three avirulent strains did not have a Ti plasmid, I built the phylogeny of 27 Ti plasmids and performed a pangenome analysis using the Roary pipeline [21]. This pipeline also generates an accessory binary genes Newick tree, which was used to make the plasmid phylogeny shown in Fig. 7.

Comparative genomics

Both NCBI prokaryotic genome annotation pipelines PGAP [22] and Prokka [23] were applied to the 29 complete genomes. The general assembly and annotation characteristic tables (Tables 3 and 4) were created using the customer Python script. The 27 Ti plasmids and 27 T6SS were manually annotated. The potential T6SS effectors with a score > 0.9 (Table 9) and scores between 0.8 and 0.9 (Table 8) were also generated using the Python script. The sequence similarity analyses of the linear chromosome, the circular chromosome, the Ti plasmids, and the At plasmids were performed using Sourmash pipeline [24].

All statistical analyses were performed in a custom R script. Other alignment graphs were generated by custom Python scripts.

Results

General genome assembly and annotation features of 29 *A. tumefaciens* strains from CG of walnut and soil samples

The genome size of 29 walnut strains ranged from 5.34 Mb to 6.31 Mb (Table 1). All strains contained at least three genetic elements: one circular chromosome, one linear chromosome, and one At plasmid (pAt). All virulent strains contained a tumor-inducing plasmid (pTi). Sixteen strains contained at least one additional replicon (Table 1), all of which were circular. The genome sizes of the circular chromosome ranged from 2.76Mb to 3.02Mb, with a G+C content of ~59%; the genome sizes of the linear chromosome ranged from 2.11Mb to 2.33Mb, with a G+C content of ~59%; the genome sizes of pTi ranged from 177kb to 252kb, with a G+C content of ~56%; and the genome sizes of pAt ranged from 312kb to 648kb, with a G+C content of ~58% (Fig. 4.5; Fig. 4.6; Table 1; Table 2).

To characterize their genome structure, the complete genomes of the 29 walnut strains were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP), and the C58 sequence was included as the annotation reference (Table 4). PGAP predicted the number of Genes (total), Genes (coding), Genes (RNA), CDSs (total), CDSs (with protein), rRNAs, complete rRNAs, tRNAs, ncRNAs, Pseudo Genes, etc (Table 4). There was a wide range in the number of total genes detected in these strains. SJ003 contained 4964 genes, whereas strain Yub001 contained 6013 genes. Most strains contained 4 rRNA operons, although But001, CL001, Sta004, and Yub002 each contained 5 rRNA operons, and Yub001 was the only strain with 6 operons. Yub001 had 62 tRNAs, whereas the remaining strains contained less than 60 tRNAs. C58 had 6 ncRNA, whereas the 29 walnut *A. tumefaciens* strains contained 4 ncRNAs. The number of pseudogenes (ambiguous residues) also varied among the 29 genomes, ranging from 37 in Tul002 to 207 in Sta001.

Sequence identity patterns of 30 *A. tumefaciens* genomes

Agrobacterim tumefaciens strains have at least three genetic elements, in addition to the pTi contained in virulent strains. To characterize the genetic diversity of each genetic element, I performed

sequence similarity analyses using the sourmash pipeline [24]. The heatmap and dendrogram of circular and linear chromosome signatures revealed a similar dendrogram pattern in terms of sequence identity based on the 51-mer profiles. Sequence signatures of strain Kin001 were the most similar to strain C58, whereas the other strains were substantially different from C58. Based on chromosomal sequence identity, strains were classified into four groups: group 1 (Tul001, SJ001, Gle001, Sta002, Sut001, Sut002, SJ002, Sta001, Sta003), group 2 (Tul004, Teh001, Teh002), group 3 (Gle002, Yol001, Yol002), group 4 (variable cluster with remaining strains) (Fig. 4.1; Fig. 4.2).

I also generated dendrograms of 29 At plasmids and 27 Ti plasmids. In the comparison of pAt sequences signatures, ten strains (Sta004, C58, Kin001, Yub002, But001, CL001, CC001, Sta005, But002, and Yub001) were significantly different from one another. The remaining 17 strains shared very similar pAt sequences signatures (Fig. 4.3). For pTi sequences signatures, the heatmap revealed three major pTi patterns. Pattern 1 contained strains Teh002, Teh001, Gle002, Sut001, and Sut002, with nearly identical sequence. Pattern 2 contained strains Kin001, Kin002, Gle001, Sta002, Tul003, and Tul004. Pattern 3 contained strains Kin003, Tul001, Yub001, Yol002, Yol001, Sta003, Sta001, SJ002, SJ001, But001, and CC001 (Fig. 4.4). Ti plasmid sequences signatures from the remaining five strains were variable.

Genetic diversity of 27 Ti plasmids

The Ti plasmid is the key genetic element that confers virulence (vir) to a given *A. tumefaciens* strain. To investigate pTi genetic diversity, I performed comparative genomic analysis on Ti plasmids from the 27 pathogenic strains (26 strains from walnut and strain C58). I manually annotated the type IV secretion system (T4SS), T-DNA region, two component regulatory systems (VirA/VirG), conjugation system, and opine catabolism genes (Table 5; Table 6; Table 7). The T4SS and T-DNA region are the two main genetic components responsible for the T-DNA transfer into the host plant cell genome. Twelve *vir* genes of the T4SS including *virB1*-*virB11* and *virD4* were successfully annotated from the 27 Ti plasmids

(Table 6). Other annotated *vir* genes include *virD1, virD2, virD3, virD5, virC1, virC2, virE1, virE2, virE3, virH1, virH2, virF* and *virK*.

The T-DNA region contains two groups of genes: i.e., hormone biosynthesis and opine synthase. The 27 Ti plasmids contained the same hormone biosynthesis genes, *gene5, iaaH, iaaM*, and *ipt*. However, the opine synthase genes were more variable, and could be divided into three groups: nopaline, agropine, and succinamopine pTi (Table 5). CL001, Sta005, and C58 contained nopaline Ti plasmids, which contained two opine synthase genes, *acs*, and *nos*. The agropine Ti plasmid was discovered in strains Gle001, Gle002, Kin001, Kin002, Sta002, Sut001, Sut002, Teh001, Teh002, Tul003, Tul004, and Yub002. Four opine synthase genes were involved in agropine biosynthesis: *acs, mas1, mas2,* and *ags*. The succinamopine Ti plasmid was present in But001, CC001, Kin003, SJ001, SJ002, Sta001, Sta003, Sta004, Tul001, Yol001, Yol002, and Yub001. These strains contained two opine synthase genes, *acs*, and *sus*.

The left-border (LB) and right-border (RB) of the T-DNA typically contain an ~25-base-pair (bp) sequences, with the RB in the wild-type orientation required for T-DNA transfer [25,26,27]. The 27 T-DNA LB and RB regions examined shared the same LB sequence, LB: 5'-tggcaggatatattgtggtgtaaac-3'. In contrast, the RBs were variable and could be divided into three types. Type I nopaline RB: 5'-TGACAGGATATATTGGCGGGTAAAC-3', present in the nopaline Ti plasmid; Type II octopine RB: 5'-TGGCAGGATATATGCGGTTGTAATT-3', present in the agropine pTi; and Type III succinamopine(sus) RB: 5'-TGACAGGATATATCAAGCTGTAGCG-3', existing in the succinamopine pTi. These data suggest that opines synthase genes and the RBs types are paired. Analysis of sequence signatures of the Ti plasmids showed that plasmids with the same opine types were clustered together. The succinamopine opine type was further divided into two subgroups based on 31-mer signatures (Fig. 4.4).

Besides the T4SS and T-DNA regions, the remaining genes on the pTi mediate opine transportation, catabolism, and conjugation (Table 5). The nopaline Ti plasmids contained *accA,B,C,D,E*; *nocP,M,Q,T* for opine transport, and *accF,G*; *noxA,B; hyuA,B* that mediate opine degradation. Agropine Ti

plasmids contained *accA, B, C, D, E*; *motD,C,B,A*; *moaB,C,D,A*; *agtC,B,A*; *agaA,C,B,D* for opine transport and *accF,G*; a*gcA*; *agaF,G*; *mocE,D,C*; *moaE* for opine degradation. Succinamopine pTi contained *accA,B,C,D,E*; *odh*; *sacD,C,B,A*; *potA,B,C* for opine transport and *accF,G*; *sacH,G,F,E*; *hyuA,B* for opine degradation. Other genes involved in opine metabolism of each Ti plasmid type included *accR; nocR; mocA,B,R,R'; moaR; sacR*.

The acc operon contains 8 genes *accR* and *accABCDEFG*, and is required for *A. tumefaciens* strains to be sensitive to biocontrol strain K84 [3]. However, though all virulent strains examined in the present study contain the acc operon, 12 of 27 strains were resistant to K84 (Fig. 4.7). Interestingly, 5 (*accBCDEF*) out of 8 acc genes of strains resistant and sensitive to K84 showed a different pattern when aligning their amino acid sequences (Fig. 4.19). Genes *accA*, *accG* and *accR* were conserved among all 27 virulent strains.

The 27 virulent strains also contained the same conjugation system including the tra and *traI/trb* regions. C58 contained two copies of the replication region, i.e., *rep*A', B', C' and *rep*A, B, C. Agropine Ti plasmids contained the same *repA', B', C'* of C58, whereas nopaline and succinamopine Ti plasmids contained the same *repA, B, C* genes as found in C58. The chemotaxis protein *mcpA* was found in 27 Ti plasmids (3 strains avirulent) (Table 5).

Pathogenicity of 30 *A. tumefaciens* strains

I performed Koch's postulates on 30 *A. tumefaciens* strains with *D. stramonium* and walnut tissue culture rootstock of hybrid (*J. hindsii* X *J. regia*) Vlach, the 30 strains showed significant differences in the fresh weight of tumors induced on *D. stramonium* and walnut used as rootstock (Chapter 3). Three strains were avirulent: Tul002, But002, and SJ003. Whole-genome assembly further revealed that these three strains lacked a Ti plasmid. Kin002 consistently generated the smallest tumors in both *D. stramonium* and walnut, whereas the largest tumor in both hosts were induced by Yol001 and Teh001, respectively. While comparing the gene components and gene sequences involved in pathogenesis, there is no obvious difference detected among the 27 T-DNA regions (Fig. 4.12, Table 6).

To investigate the relationship between pathogenicity and RB DNA sequence, I aligned the phylogeny of 27 Ti plasmids, the RBs, and the phenotypic traits (Fig. 4.7). Depending on opine type, I found that the tumor weights on *D. stramonium* and walnut were positively correlated (Fig. 4.15). Interestingly, the tumor weights of sus Ti plasmids were generally larger than that induced by agropine type Ti plasmids in *D. stramonium* (Fig. 4.8). In walnut, the tumors caused by agropine type pTi were generally larger than those caused by succinamopine pTi containing strains (Fig. 4.9).

The opine type had a statistically significant impact on *D. stramonium* tumor size with a p-value of 0.006 (Fig. 4.8), whereas the impact on the walnut tumor was not significant with a p-value of 0.05889 at alpha = 0.05 of the t-test (Fig. 4.9). To study the impact of host type on tumor weight, I performed a t-test on *D. stramonium* and walnut tumors induced by the same opine type. Interestingly, plant host had a significant impact on tumor weight caused by either agropine Ti plasmids (p-value = 0.004) (Fig. 4.10) or by succinamopine Ti plasmids (p-value = 0.05) (Fig. 4.11).

T6SS of 29 *A. tumefaciens* strains

To investigate the genes that comprise the T6SS, I performed a BLAST analysis of the 29 *A. tumefaciens* linear chromosomes of walnut strains against the T6SS of *A. tumefaciens* C58. Twenty-six out of twenty-nine walnut strains contained the T6SS. Three virulent strains (CC001, Tul003, and Yub001) did not appear to have a T6SS. Interestingly, Yub001 is the fastest growing strain in rich medium TSB and slowest in poor medium AB. The 27 T6SSs (26 walnut isolates plus C58) consisted of two operons, the *imp* and the *hcp* operons, which were oriented in the opposite direction on the linear chromosome. The *imp* operons were conserved in sequence identity and their structure, ranging from *tagE*[1] to *tssA*[14]. However, the *hcp* operons were variable and were classified into eight groups, named as *hcp*_c1 to *hcp*_c8 (Fig. 4.13). Genes *tssH* and *tssD* were present in all 27 *hcp* operons. Tai-tae effector-immunity (EI) pairs were present in 26 strains but were absent in strain CL001. Gene *tssI-1* was present in 26 strains, whereas Kin002 contained one copy of *tssI-1* located in the region [1555511:1557877] of the linear chromosome.

The EI pair *tde-tdi* EI pair was only found in strains C58 and Kin001. This EI pair was followed by *paaR*, which was present in four *hcp* groups, *hcp*_c1, *hcp*_c2, *hcp*_c4, and *hcp*_c7. Genes upstream of *paaR* were functionally unknown in these four groups. Among the eight *hcp* groups, *hcp*_c1, *hcp*_c2, and *hcp*_c3 contained a DUF4123 domain-containing protein. Hcp_c5 and hcp_c8 exhibited a DUF2169 domain-containing protein and a DUF4150 domain-containing protein.

C58 contained a second copy of *tss*I-2 gene, encoding the VgrG2 (atu3642-atu3636) operon encoding the VgrG module. Genomic comparison of the VgrG module revealed that 18 strains contained a second copy of the *tssI-2* gene, whereas eight strains did not (Fig. 4.14). After examining genes downstream of the *tss*I-2, I found 15 strains have the same structure in terms of the genes synteny, whereas the other three were variable. Similar to the C58 VgrG2 module, 15 strains exhibited a gene encoding a predicted DUF2169 domain-containing protein following the tssI-2 protein. Interestingly, the hcp operon in these 15 strains contained a gene predicted to express a DUF4123 domain-containing protein (Fig. 4.13).

To comprehensively investigate T6SS effectors, I selected 15 linear chromosomes based on their sequence identity (Fig. 4.2) for T6SS effectors prediction using the Bastion6 tool. The prediction results showed that each of these 15 strains contained hundreds of T6SS putative effectors. Only putative effectors with score > 0.8 were included for further analysis. First, I selected putative effectors with a score > 0.9. Eight effectors were identified. Among them, *Tae*, *tde*1, *tss*D, and *tss*I have been functionally characterized in *A. tumefaciens* [28]. Another 4 putative effectors, including a PhoX family protein, a RHS repeat protein, glycosyl hydrolase, and one hypothetical protein in *A. tumefaciens* (Table 9).

Secondly, using a cut-off score of 0.8, 40 putative T6SS effectors were identified in the 15 strains. Among them, 12 were hypothetical proteins. The rest were a DUF4962 domain-containing protein, a DUF4038 domain-containing protein, a DUF4150 domain-containing protein, polygalacturonase, vitamin_B12_transporter BtuB, SGNH/GDSL hydrolase family protein,

outer_membrane_beta_barrel_protein, glycosyl hydrolase, colicin I receptor, Endo-1,3-1,4-beta-glycanase ExoK, 5-deoxy-glucuronate isomerase, catalase-peroxidase, *tss*I type VI secretion system tip protein VgrG, DKNYY domain-containing protein, malto-oligosyltrehalose trehalohydrolase, alginate lyase, exo-alpha-sialidase, iron ABC transporter permease, glycogen operon protein GlgX, glycoside_hydrolase_family_28_protein, cupin, putative_deoxyribonuclease RhsC, RHS domain-containing protein, Glucoamylase, tae4, HNH endonuclease, HNH endonuclease, and a actin cross-linking toxin VgrG1. Most predicted putative T6SS effectors had a score < 0.8.

To investigate how putative T6SS effectors correlated with the phenotypic data collected in Chapter 3, I performed PCA and Pearson correlation analyses on the number of T6SS effectors of each strain (dft6es), *D. stramonium* tumor sizes (tumor_means), walnut tumor size (walnuttumor_means), K84 sensitivity (K84_mean), the growth rate in rich TSB medium (tsb_mean), and growth rate in poor AB medium (ab_mean). Pearson correlation analysis and PCA results both revealed that the mean size of *D. stramonium* tumor and walnut tumor were positively correlated (correlation coefficient of 0.67, Fig. 4.15). No correlation was observed among the number of T6SS effectors, the tumor size and K84 sensitivity. However, PCA analysis showed that the number of T6SS effectors was negatively correlated with the growth rate in nutrient-poor medium, but positively correlated with growth in the nutrient rich medium TSB (Fig. 4.15).

Besides virulence and growth rate, I examined K84 sensitivity for the 30 strains. *A. radiobacter* strain K84 is antagonistic to pathogenic *A. tumefaciens* strains by secreting the bacteriocin Agrocin84. Fifteen strains of the 30 strains were resistant to Agrocin84, and 15 strains were sensitive. All 12 succinamopine strains were sensitive to K84, and nearly all 12 agropine strains were resistant to it. Strain Tul003 was exceptional. The three avirulent strains But002, SJ003, and Tul002 were also resistant to K84 (Fig. 4.8).

All 30 strains were resistant to streptomycin (Chapter 3). Genomic comparison revealed one locus homologous to the streptomycin resistance gene *aph(6)-Id* in the circular chromosomes of all 30 strains. The *aph(6)-Id* was characterized from *Salmonella enteritidis* [29]. By aligning the 30 *aph(6)-Id* genes, I found that strains Sta005, But002, Gle001, SJ001, SJ002, Sta001, Sta002, Sta003, Sut001, Sut002, Teh001, Teh002, Tul001, Tul004 exhibited the identical amino acid sequence. However, there was great diversity in amino acid sequences of the same gene in strains Yub001, CL001, Sta004, But001, Yub002, C58, Kin001, Kin002, Tul003, SJ003, Kin003, Tul002, CC001, Gle002, Yol001, Yol002.

Pangenome analysis of 27 Ti plasmids

The 29 *A. tumefaciens* strains from walnut orchards were resolved into eight ANI groups and six genomic species: G1, G2, G4, G7, G8, G13. In addition, I identified two novel genomic species (Chapter 1). Most strains (58.6%) from walnut orchards belonged to genomospecies G4. Three strains are non-virulent. Therefore, 27 Ti plasmids including C58 were used for pangenome analysis.

The pangenome of 27 Ti plasmids contained 489 genes. Of these, only 76 were core genes. The number of shell genes and cloud genes were 252 and 161, respectively (Fig. 4.16). Phylogenetic profile combined with the gene presence/absence data demonstrated that the 27 Ti plasmids could be classified into three groups (Fig. 4.17). These three groups matched the three Ti plasmid types based on their RB and opine types (Fig. 4.12). Ti plasmids of strains Sta005, C58, and CL001 were the most variable and they had a nopaline-type Ti plasmid. The remaining strains contained either agropine or succinamopine type Ti plasmid. Interestingly, strains containing succinamopine Ti plasmids could be further divided into two subgroups, indicating great diversity in this pTi type (Fig. 4.17). Virulent *A. tumefaciens* strains from walnut orchards in California mainly harbored either an agropine (12/25 strains) or a succinamopine pTi (12/25 strains) [30].

Discussion

Understanding the genetics, biology, and ecology of plant pathogens facilitates development of plant disease management strategies in agricultural ecosystems. Crown gall has been a chronic disease in California walnut orchards since walnut cultivation started in the late 1700s [31]. Currently, about 85% of the walnut industry uses the Paradox rootstock, a hybrid of *Juglans hindsii* X *J. regia* [32], which unfortunately is highly susceptible to *A. tumefaciens*. Exacerbating this problem is a lack of information on both the genetic and phenotypic diversity of *A. tumefaciens* strains occurring in the walnut growing region of California. Characterization of *A. tumefaciens* diversity will facilitate development of sustainable pre- and post-plant CG management strategies, including the identification of CG-resistant rootstock genotypes. Here, I present a case study on the genetic diversity of *A. tumefaciens* strains collected from the top ten walnut growing counties of California, which will contribute to filling this knowledge gap.

In this dissertation I examined *A. tumefaciens* genomic diversity, and the diversity of five phenotypes, i.e., virulence on two hosts, antibiotic resistance, K84 sensitivity, and growth rate in both nutrient-rich and nutrient-poor media (Chapter 3). An unexpected level of phenotypic and genetic diversity was observed in the 29 *A. tumefaciens* strains from walnut orchards examined (Chapter 3). Our observation of multiple strains with genetically diverse chromosomes harboring genetically similar Ti-plasmids is consistent with observations by Weisberg et al. (2020), which support the idea that there exist disease reservoirs in this agricultural ecosystem [33]. However, examination of 143 oncogenic plasmids by Weiseberg, et al. (2020) revealed six types of Ti plasmids and three types of Ri plasmids [33]. Using this classification, the nopaline Ti plasmids in our collection belong to Type I.a pTi, whereas the agropine and the succinamopine Ti plasmids belong to Type III pTi.

I did not detect a correlation between geographical location (10 counties) and the phylogeny of 27 Ti-plasmids (Fig. 4.7). This indicates that the *A. tumefaciens* strains were most likely introduced into various California walnut orchards from multiple sources and is consistent with operations of the nursery

trade in California where a few major nurseries provide trees to many different walnut growing regions acrossthe state. This may occur if asymptomatic trees harbor rhizosphere populations of a given *A. tumefaciens* genotype which may be introduced across a wide geographic area.

Although the agropine and succinamopine Ti plasmid belong to the Type III pTi, I found members of this group to be extremely diverse, both genetically and phenotypically. Strains containing a succinamopine pTi also appear to be more virulent on *D. stramonium*, while strains harboring an agropine pTi appear to be more virulent to walnut. If the notion that different opine types confer a host-preference, this would explain why the same Ti plasmids have a significant impact on *D. stramonium* tumor size and little impact on tumor size on walnut. There may exist co-evolution among strains of different opine types and the walnut host over the decades after strain introduction into walnut orchards. This long-term interaction may have led to host adaptation resulting in a more limited impact on walnut tumor size.

Our comparative genomics examination also identified great diversity among the genetic elements mediating K84 resistance and growth rates. *Agrobacterium tumefaciens* strains that harbor agrocinopine metabolism genes are known to be sensitive to K84, due to agrocin 84 mimicking agrocinopine and inhibiting RNA and DNA synthesis [34]. However, all 27 virulent strains contain the agrocinopine metabolism genes, yet only 12 were resistant to K84. Others have shown that strains with mutations in *accF* could take up agrocin 84, but remained resistant to the antibiotic [3]. Interestingly, almost all strains with agrocinopine Ti plasmids were resistant to K84. Two different SNP patterns observed in the *accBCDEF* genes in the acc operon of different opine Ti plasmids may explain why the 12 virulent strains were resistant to K84.

I also identified potential genetic elements that mediate or are associated with growth rate. The growth rate hypothesis described by Elser *et al* (2000) suggested that the increased copy number of rRNA genes was associated with faster growth rates [35]. Interestingly, the fastest growing strain in TSB was Yub001, which contains the greatest number of rRNA genes (i.e., 6 rRNA genes), whereas all the slower

growing strains contained either 4 or 5 copies of rRNA genes. Interestingly, it has been reported that starvation or limitation of growth by nitrogen deficiency will facilitate a decoupling of bacterial growth with RNA allocation and P content [36]. Our results support these observations. That is, the fast-growing strains Yub001, But001, and But002 in nutrient-rich medium have the slowest growth rate in nutrient-poor AB medium.

PCA analysis revealed that the growth rate in AB medium was negatively correlated with the number of T6SS effectors. Also, strain Kin002 induced the smallest tumors on both *D. stramonium* and walnut, while it has the greatest number of predicted effectors with a total of 19. Effectors are bacterial "weapons" secreted to confer a fitness advantage to donor strains, and are often inhibitory to adjacent other microbes. The immunity protein in the effector-immunity (EI) pair protects the donor strain from its own toxin [37]. Secreting effectors has a metabolic cost for *A. tumefaciens,* which may lead to a slower growth rate in the nutrient poor AB medium. Use of a T6SS and ability to initiate tumor formation are two survival strategies utilized by *A. tumefaciens*. However, there is likely an ecological and/or metabolic cost for selecting one strategy over the other. Secreting a plethora of effectors leads to a reduced growth rate in nutrient-poor medium, whereas reducing competition from the adjacent microbial community.

I detected three virulent strains which lacked a T6SS. This supports the concept that T6SS's are not required for plant pathogenesis, but may be important for survival and competition [18]. The conserved structure of the *imp* operon and variation in the *hcp* operon of the T6SSs are consistent with the observations reported by others for *A. tumefaciens* strains [38]. Examination of 11 *A. tumefaciens* strains representing the 4 genomic species G1, G4, G7 and G8 revealed 6 diverse *hcp* operons[17],[38]. The authors also showed that the T6SS expression varied as a function of in-planta environment, whereas facilitating adaptation to various niches [18]. Here, I report eight groups of the *hcp* operon among the 30 *A. tumefaciens* strains examined (Fig. 4.13). Given our collection is from walnut orchards in California, this observed T6SS diversity indicates *A. tumefaciens* was most likely introduced to walnut orchards from multiple sources

over a long period of time, which is consistent with multiple nursery sources supplying trees to similar locations. The gene pair in *hcp*_c1 operon located upstream of the gene *paaR* contains one gene encoding GNAT family N-acetyltransferase. In the same region of C58 [16] and Kin001 of *hcp*_c2, they are *tde-tdi* EI pair [39]. Considering the highly identical sequence and structure of other regions of the *hcp* operon, it is reasonable to consider that the gene pair in *hcp*_c1 may serve as an EI pair. EI pairs of *A. tumefaciens* strains are either DUF4123-associated or DUF2169 associated proteins [18]. These are also the primary pairs detected in our data, except in *hcp*_c4 and *hcp*_c8, which are functionally unknown proteins.

Our research provides insight into the genetic diversity of *A. tumefaciens* strains from a defined geographical location isolated from the same host. In this effort, I focused on a variety of phenotypes that contribute to survival strategies used by this pathogen and identified potential genetic elements that correlated with the observed phenotypic differences. It is essential for breeders to consider the genetic diversity data when breeding CG-resistant walnut rootstocks in the future. This information also provides new insights into developing an expanded toolkit for the management of CG. Use of antibiotics in agriculture should be strictly regulated given that the streptomycin resistance gene is found in all our strains, which is one of the most broadly used antibiotics in controlling fire blight of apple caused by *Erwinia amylovora* [11]. Selecting additional biocontrol strains from nature, (i.e., both new bacteriocin producing *A. tumefaciens* strains and other species) and using a mix of these strains will enhance our strategies to manage CG of woody perennials.

Figures and Tables

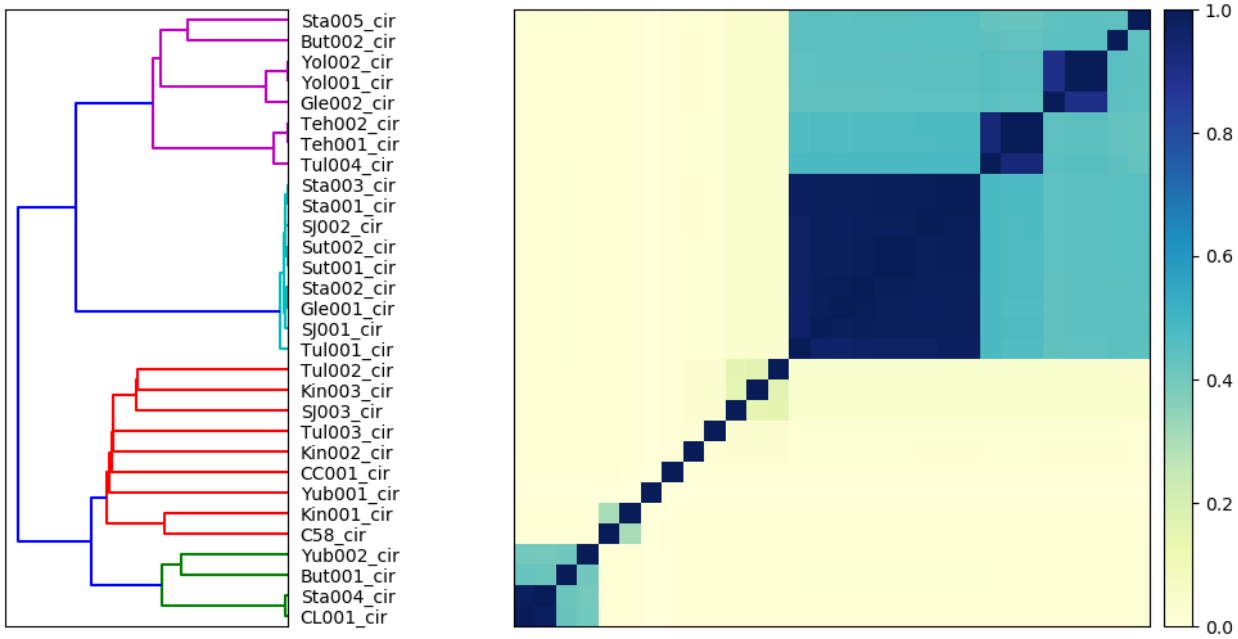


Fig. 4.1. Heatmap and dendrogram generated using sourmash signature built from the entire circular chromosome DNA sequences of 30 *A. tumefaciens* strains. Strain designations indicated vertically on right of the dendrogram. The color changes from yellow to blue indicate sequence similarity from 0 (yellow) to 100% (dark blue).

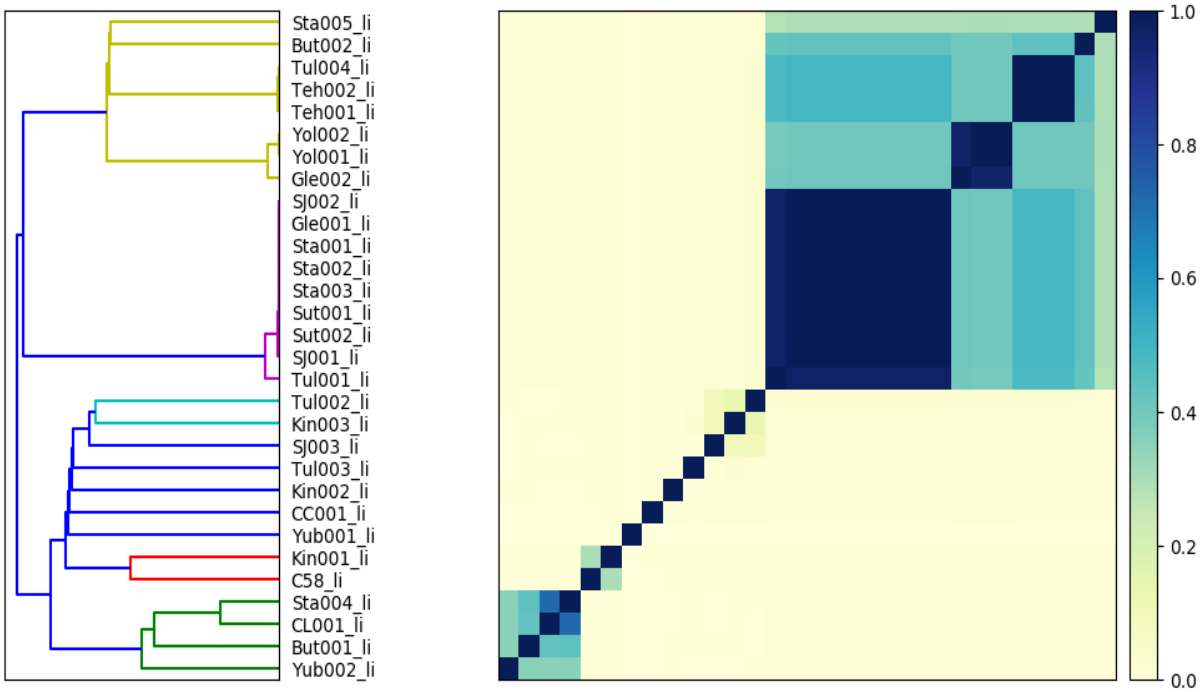


Fig. 4.2. Heatmap and dendrogram generated using sourmash signature built from the entire linear chromosome sequences of 30 *A. tumefaciens* strains. Strain designations indicated vertically on right of the dendrogram. The color changes from yellow to blue indicate sequence similarity from 0 (yellow) to 100% (dark blue).

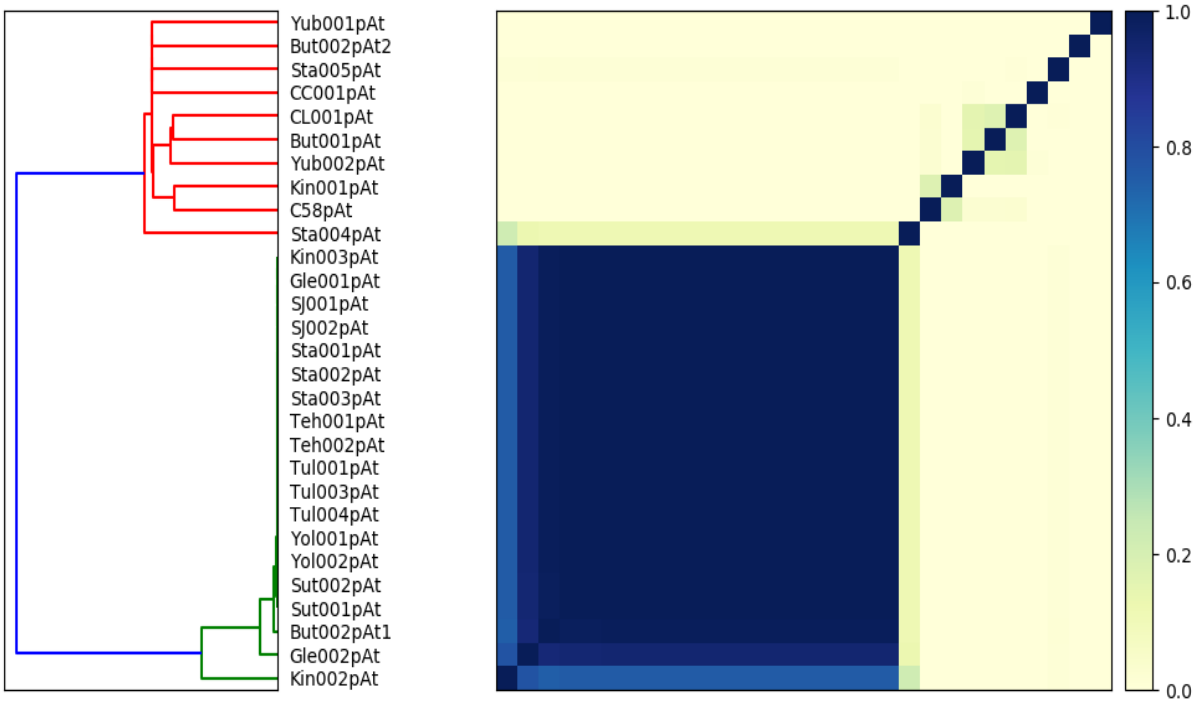


Fig. 4.3. Heatmap and dendrogram generated using sourmash signature built from At plasmids sequences of 30 *A. tumefaciens* strains. Strain designations indicated vertically on right of the dendrogram. The color changes from yellow to blue indicate sequence similarity from 0 (yellow) to 100% (dark blue).

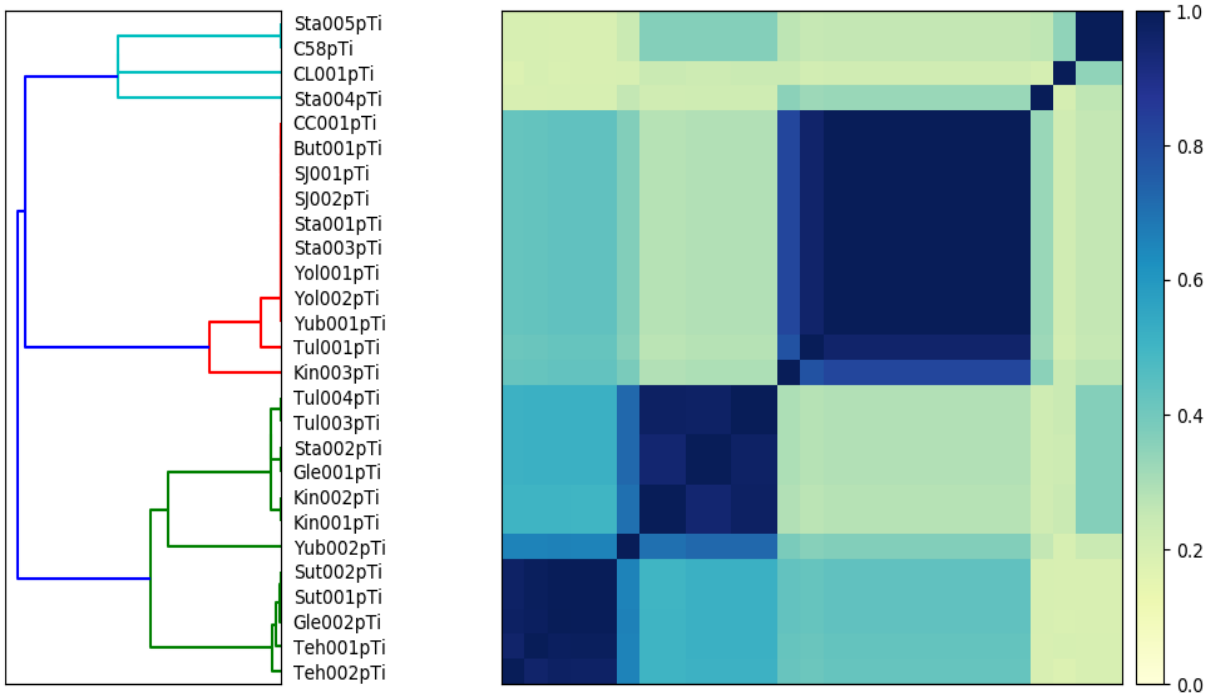


Fig. 4.4. Heatmap and dendrogram generated using sourmash signature built from tumor-inducing plasmids sequences of 27 virulent *A. tumefaciens* strains. Strain designations indicated vertically on right of the dendrogram. The color changes from yellow to blue indicate sequence similarity from 0 (yellow) to 100% (dark blue).

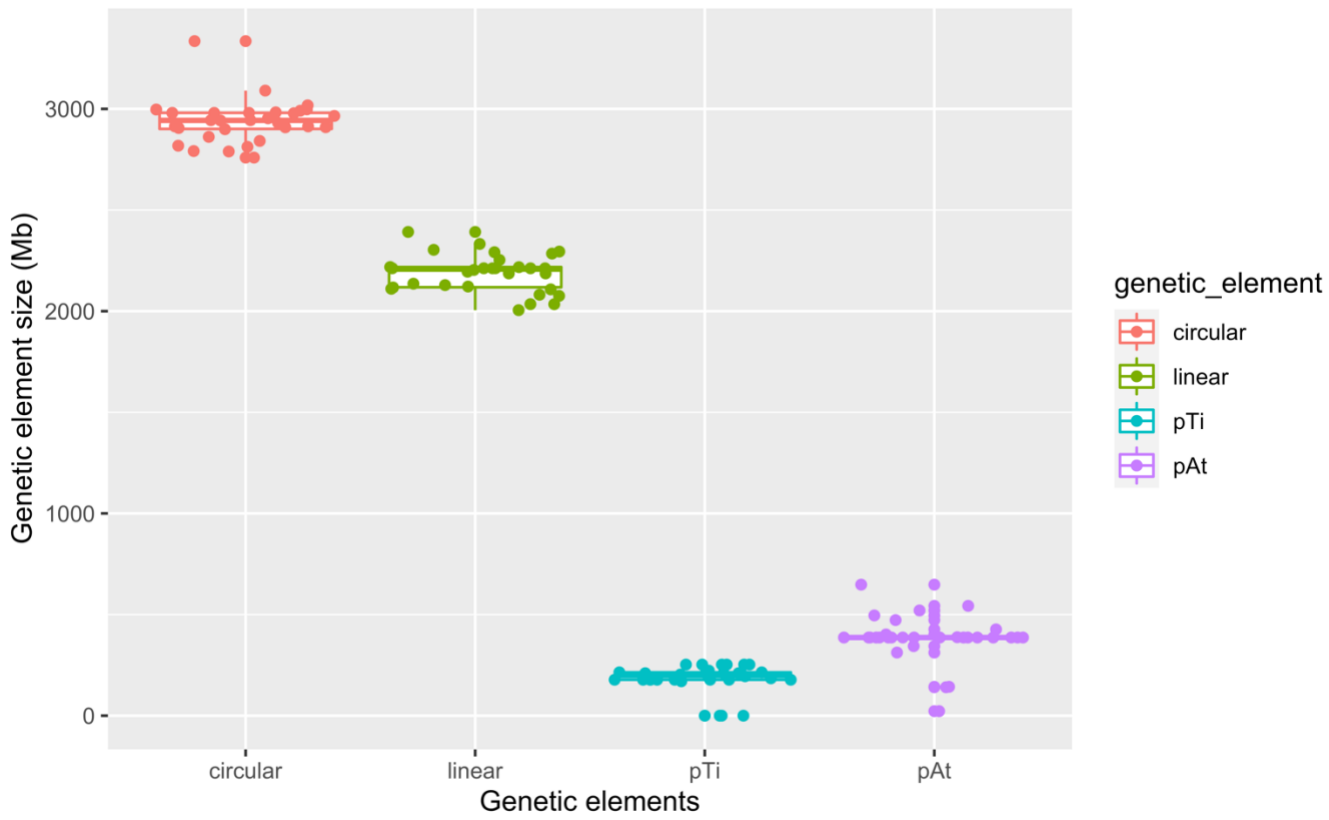


Fig. 4.5. Boxplot of the genome size of each genetic element of the 30 *A. tumefaciens* strains, including circular chromosome, linear chromosome, At plasmids, and Ti plasmids. The X-axis are the four genetic elements. Y-axis is the genetic element size in Mb.

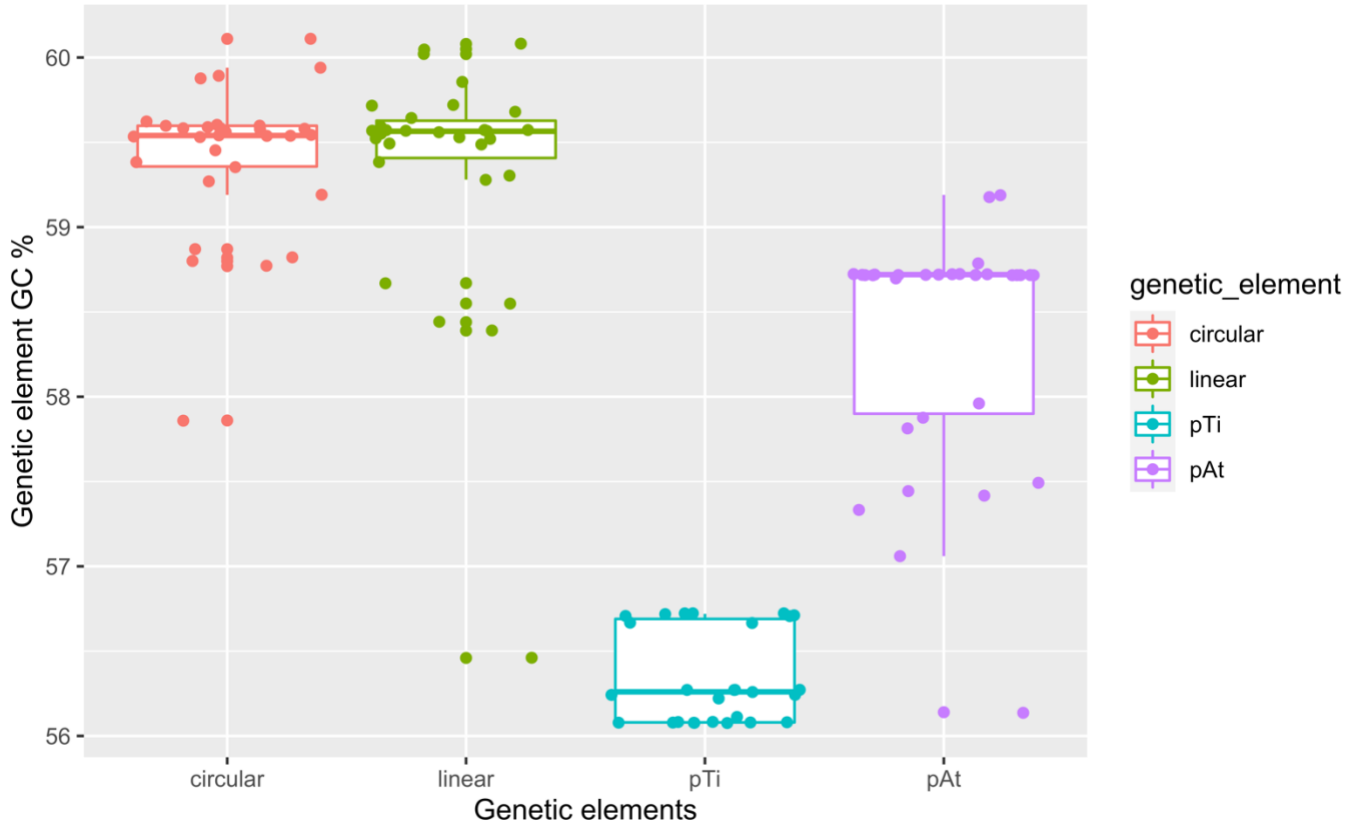


Fig. 4.6. Boxplot of the GC content each genetic element of the 30 *A. tumefaciens* strains, including circular chromosome, linear chromosome, At plasmids, and Ti plasmids. The X-axis are the four genetic elements. Y-axis is the GC content of each genetic element.

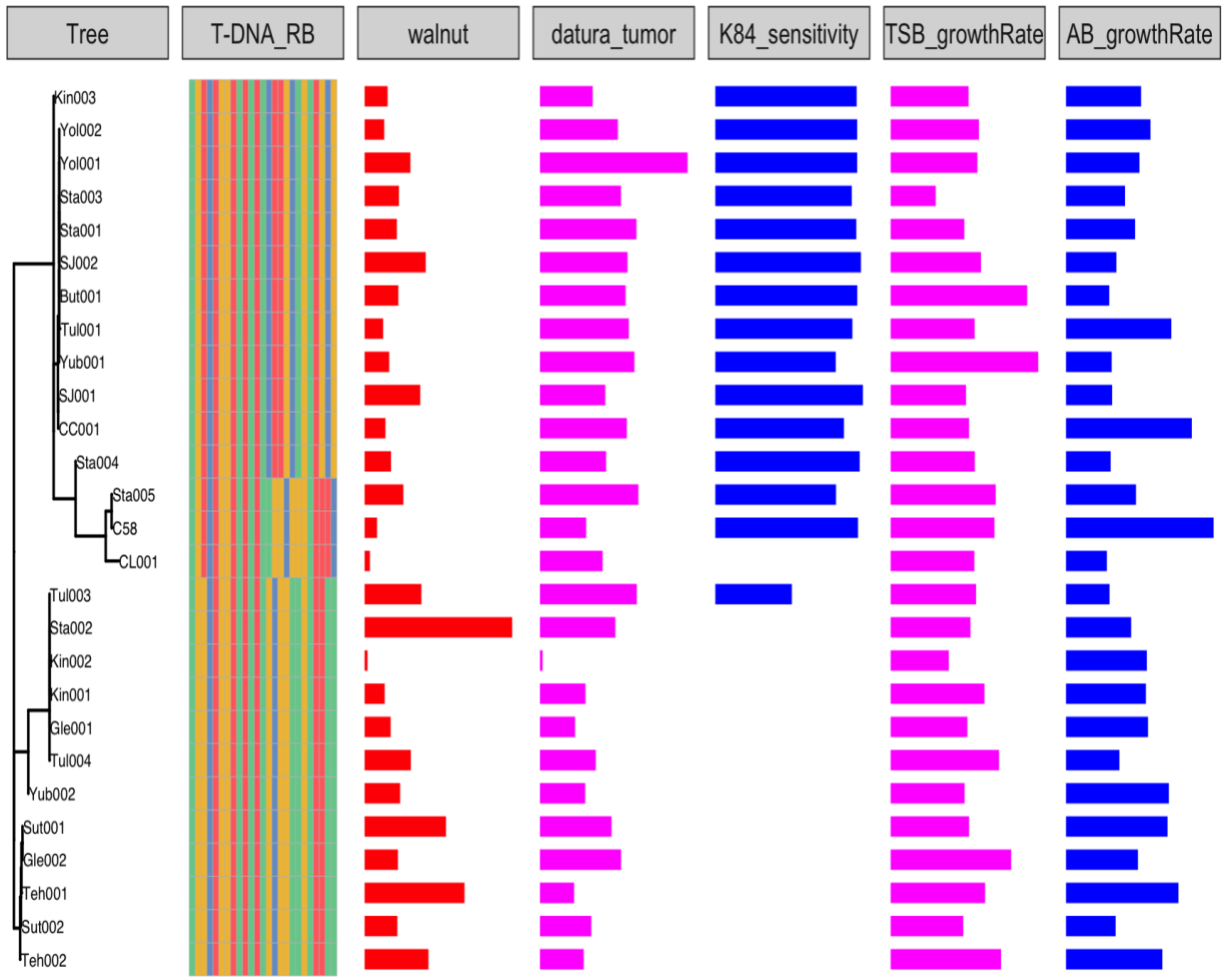


Fig. 4.7. pTi phylogeny aligned with T-DNA right border, walnut=walnut tumor size, Datura=Dautura tumor size and the inhibition zone of K84 sensitivity test, growth rate in TSB medium, growth rate in AB medium. *A. tumefaciens* strains are indicated vertically to the right of the pTi phylogeny. The four colors in the T-DNA_RB represents the four nucleotides: A(red), T(green), C(blue), G(yellow). Each bar following the T-DNA_RB represents each phenotypic trait measured in Chapter 3. The longer bar length corresponds the bigger value, (i.e, in Bar K84_sensitivity, each bar length is the diameter value of the inhibition zone of each train growing on the K84 medium. The missing bar represents the corresponding value is 0, which indicates those strains are resistant to K84.

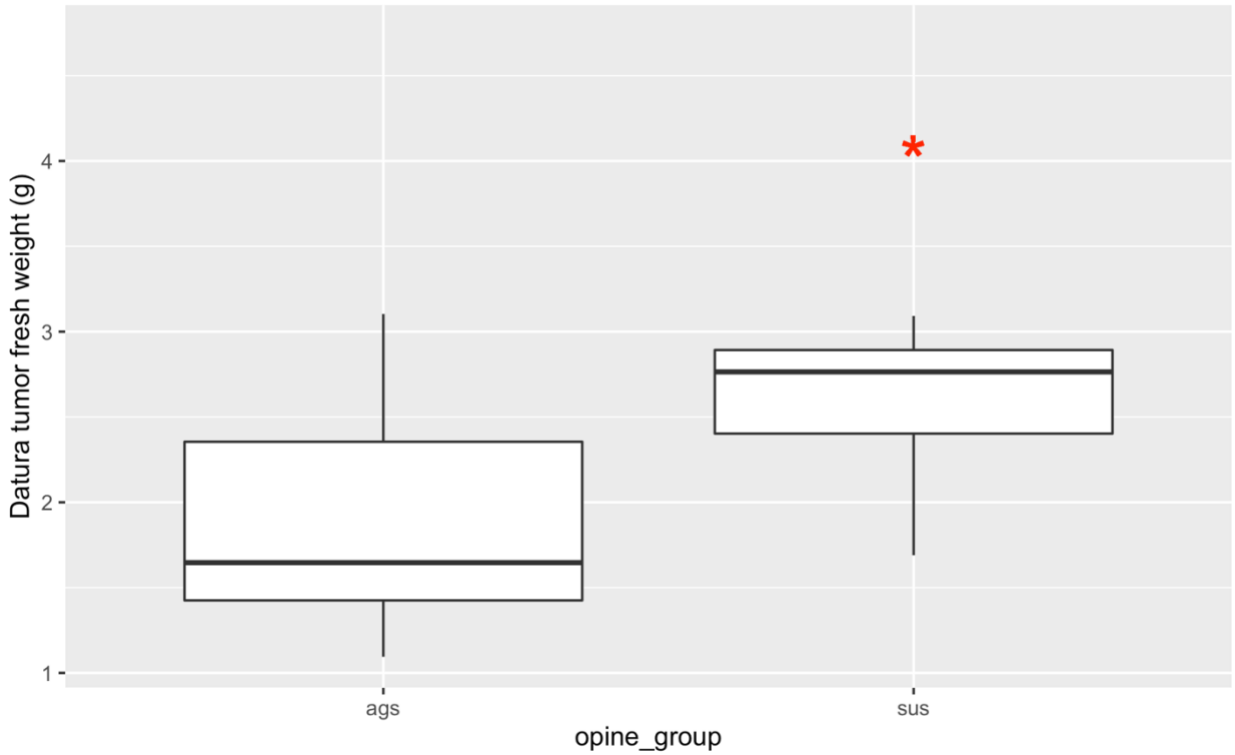


Fig. 4.8. Boxplot of the Datura tumor fresh weight caused by agropine Ti plasmid and succinamopine Ti plasmid containing *A. tumefaciens* strains. The X-axis are the two opine types: ags-agropine type *A. tumefaciens* strains; sus-succinamopine *type A. tumefaciens*. T-test is applied at the 95% confidence level.

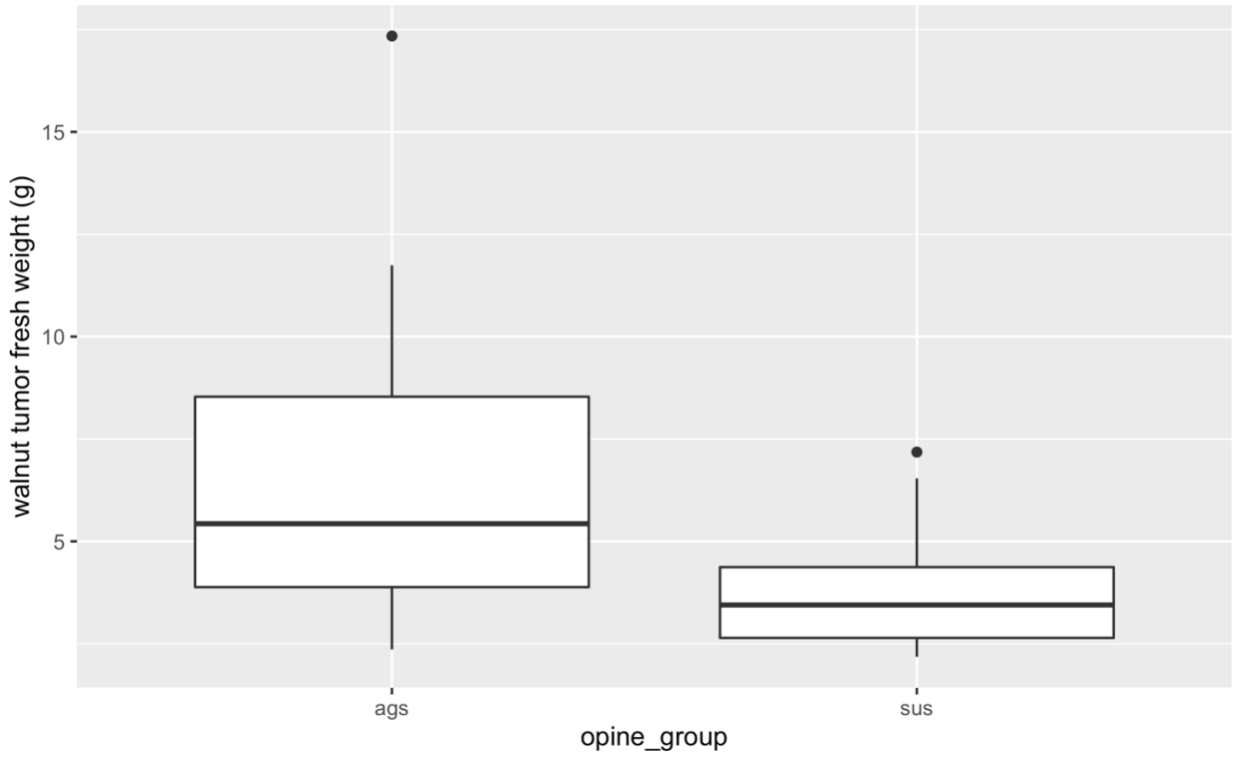


Fig. 4.9. Boxplot of the walnut tumor fresh weight caused by agropine Ti plasmid and succinamopine Ti plasmid containing *A. tumefaciens* strains. The X-axis are the two opine types: ags-agropine type *A. tumefaciens* strains; sus-succinamopine type *A. tumefaciens*. T-test is applied at the 95% confidence level.

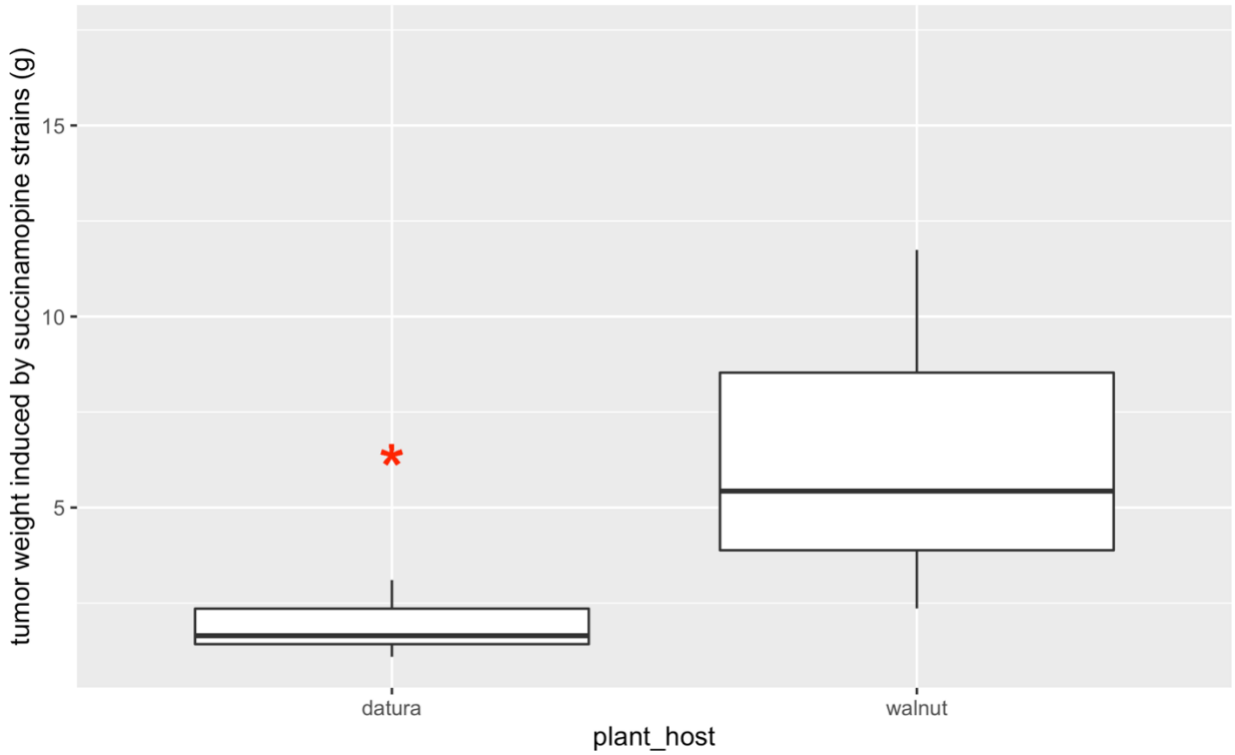


Fig. 4.10. Boxplot of the tumor weights from *Datura stramonium* and walnut plants induced by agropine Ti plasmids. The X-axis represents the two plant hosts: Datura and walnut. The Y-axis is the tumor fresh weight induced by agropine strains on the two hosts. T-test is performed at the 95% confidence level.

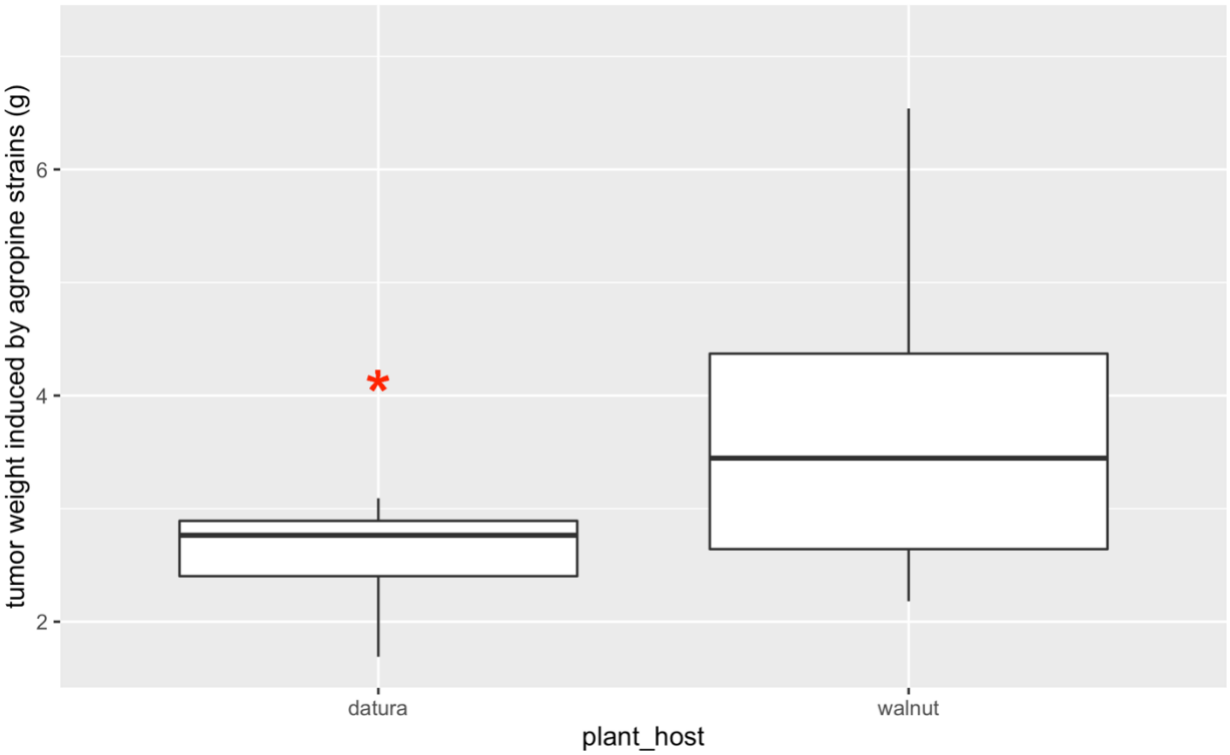


Fig. 4.11. Boxplot of the tumor weights from *Datura stramonium* and walnut plants induced by succinamopine Ti plasmids. The X-axis represents the two plant hosts: Datura and walnut. The Y-axis is the tumor fresh weight induced by succinamopine strains on the two hosts. T-test is performed at the 95% confidence level.

.

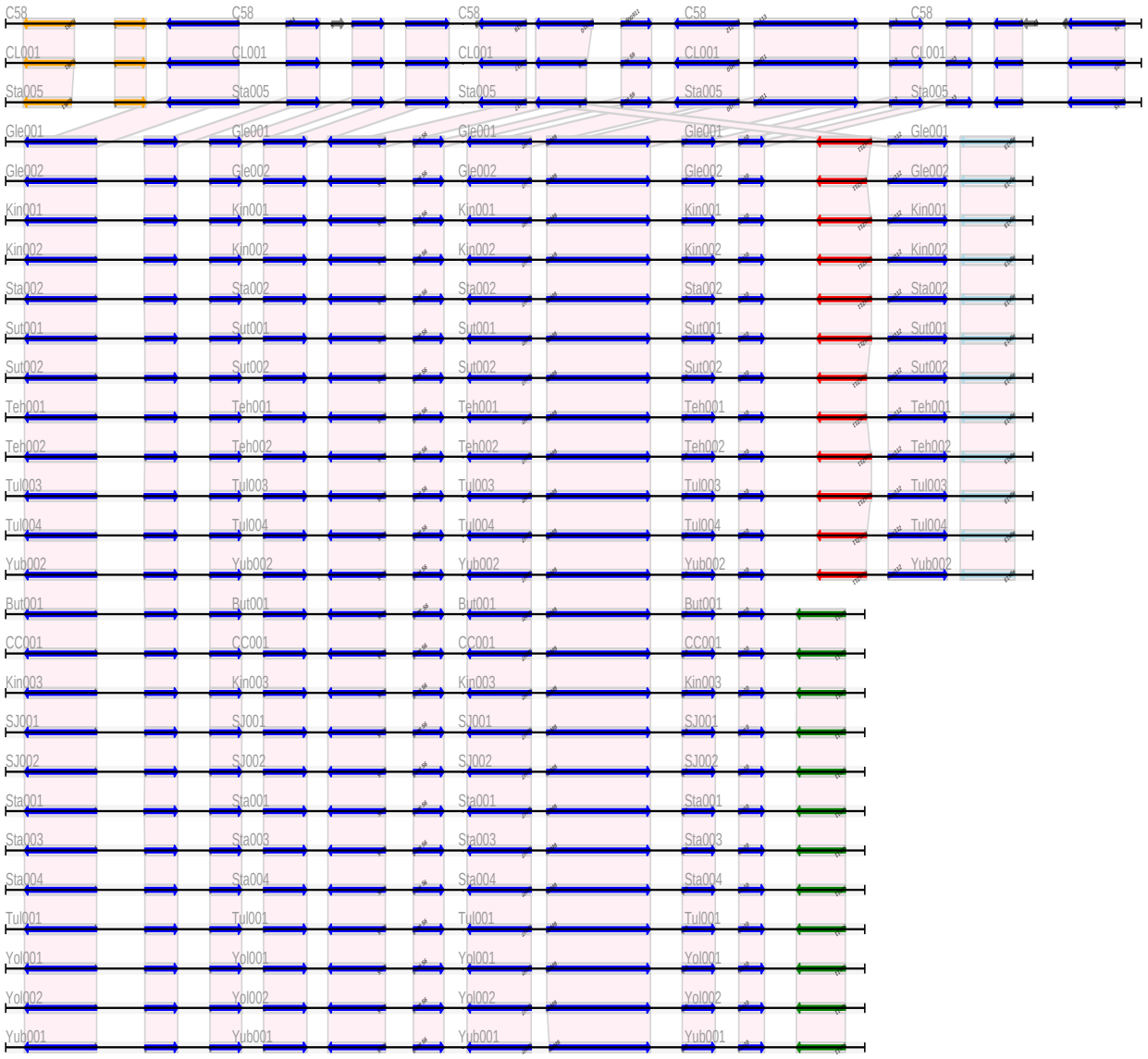


Fig. 4.12. The alignment of the T-DNA regions of Ti plasmids from 27 *A. tumefaciens* srains isolated from the top ten walnut growing counties in California. Each arrow represents the coding region of one gene (CDS). The arrow point represents the gene orientation. Each gene name is marked at the starting point of each arrow. The space between two genes is the intergenic space. The pink color is the cross-links between genes indicating their similarity. The arrows with the same color showed the conserved genes among different T-DNA regions. Each strain name is repeatedly marked on each alignment, i.e., C58.

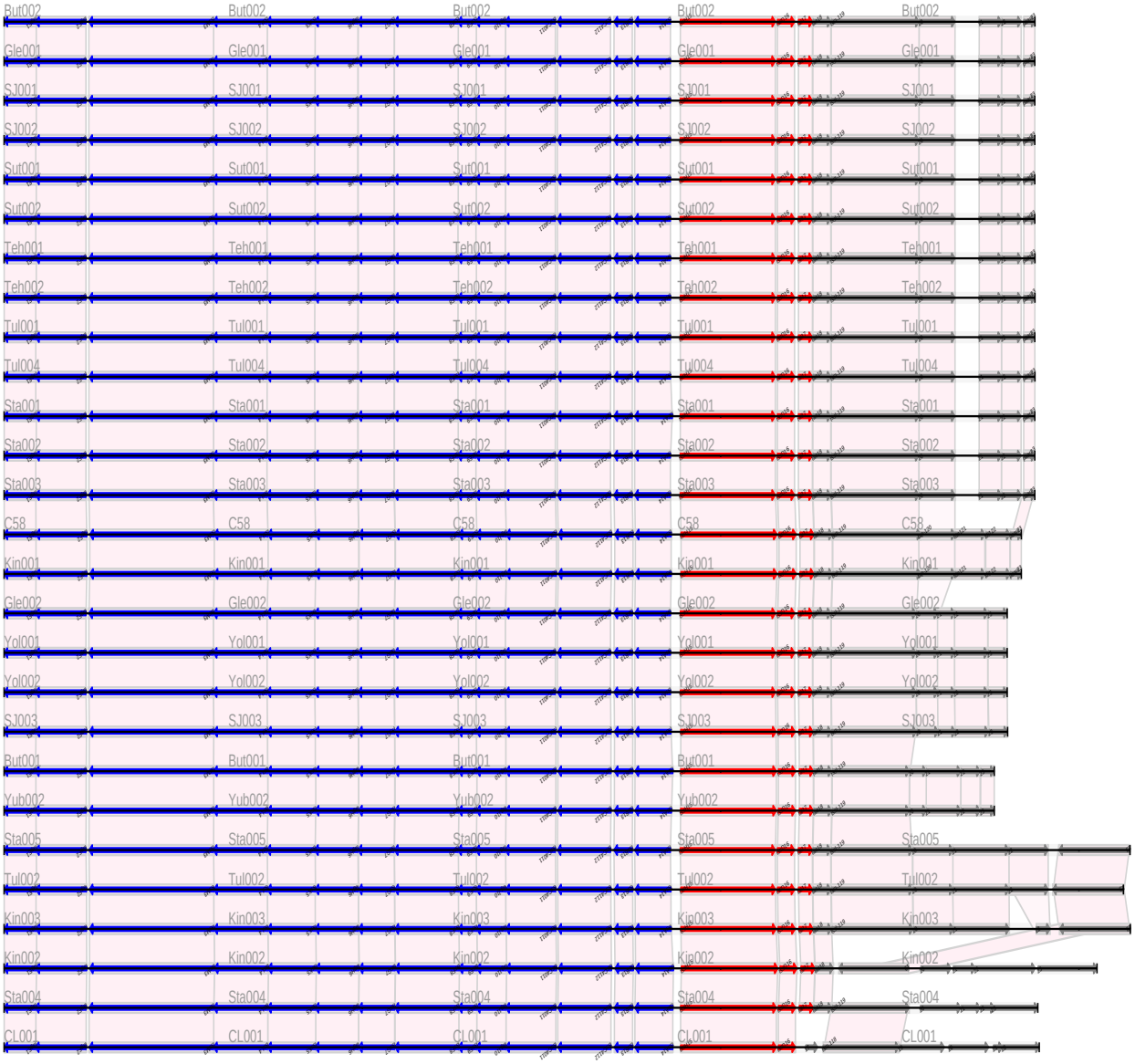


Fig. 4.13. The alignments of genes of the T6SS of 27 *A. tumefaciens* strains. The space between the blue and the red arrow is the boundary of two operons: the imp operon marked in blue and the hcp operon consisting of genes marked in red and its downstream genes. Each arrow represents one gene, and the gene name is marked at the start point of the gene. The pink color is the cross-links between genes indicating their similarity. The arrows with the same color showed the conserved genes in the different T6SS. Each strain name is repeatedly marked on each alignment, i.e., C58.

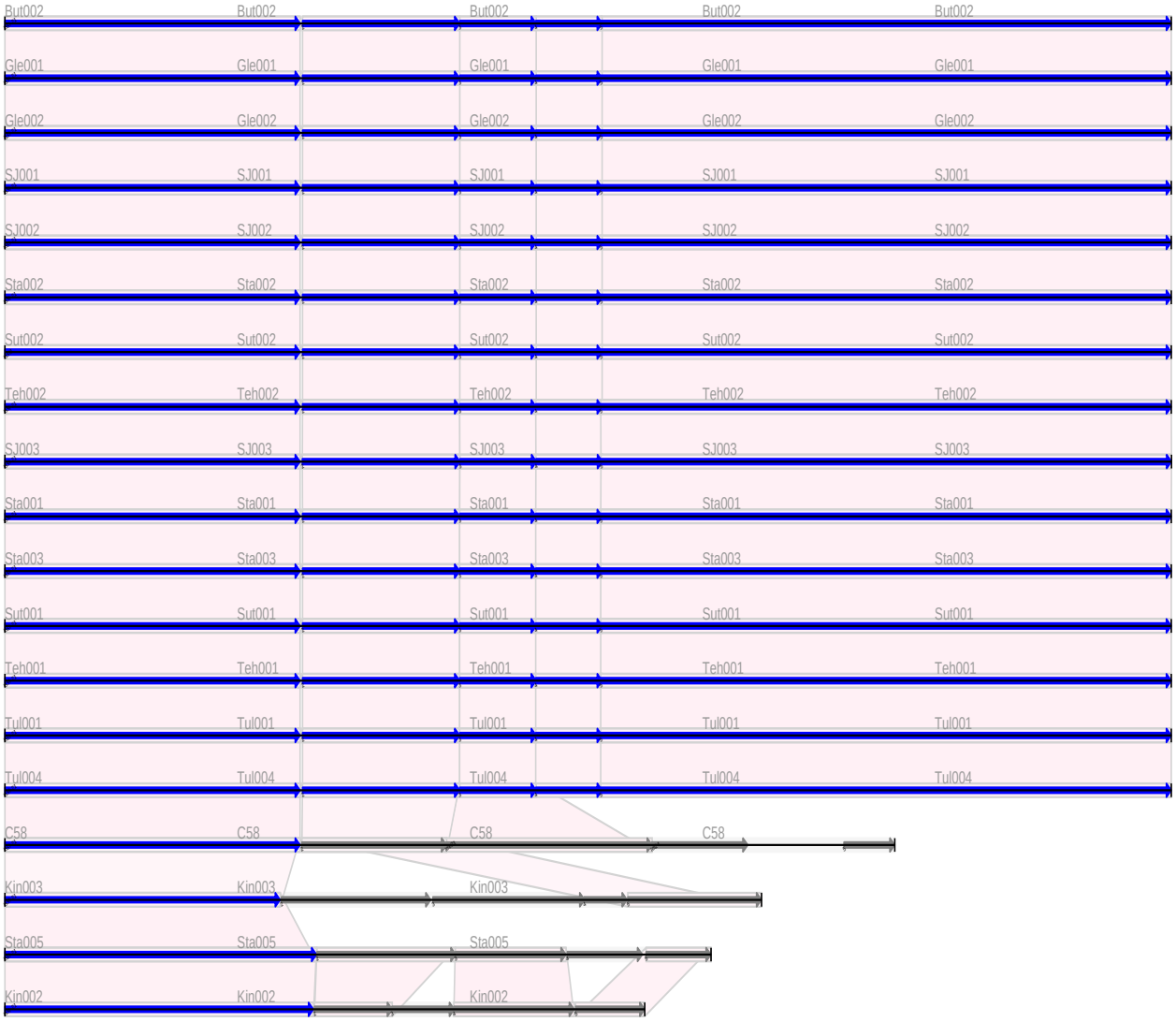
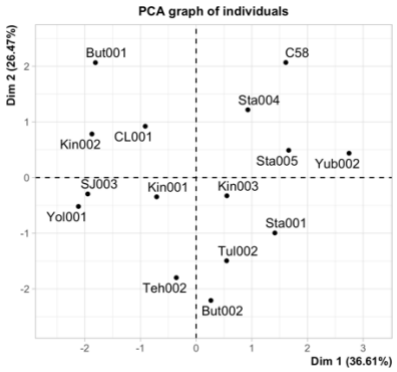


Fig. 4.14. The alignment of the VgrG2 module of 15 *A. tumefaciens* strains isolated from the top 10 walnut growing regions of California. Each arrow represents one gene. The pink cross links show the conserved genes among different strains. Each strain name is repeatedly marked on each alignment, i.e., C58.

A.



B.

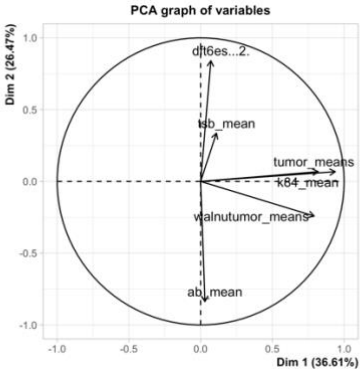


Fig. 4.15. PCA analyses of the number of predicted T6SS effectors, the growth rate in TSB medium, the growth rate in AB medium, datura tumor weight, walnut tumor weight, and K84 test. Left: PCA graph of individual strains in two dimensions. Dim 1 was 36.61 % and Dim 2 was 26.47%. Right: PCA graph of variables in two dimensions with Dim 1 36.61% and Dim 2 26.47%. The growth rate in AB medium were negatively correlated with the number of potential T6SS effectors. The growth rate in TSB medium was positively correlated with the number of predicted T6SS effectors. The walnut tumor weight and the datura tumor weight were positively correlated.

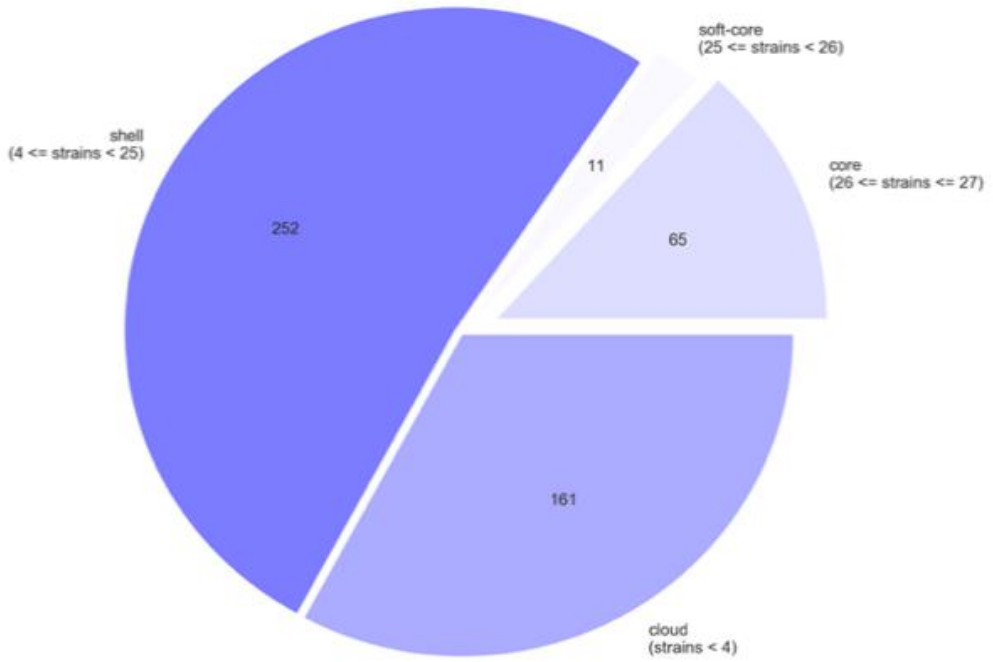


Fig. 4.16. The pangenome of 27 Ti plasmids isolated from 27 virulent *A. tumefaciens* strains isolated from the top 10 walnut growing counties in California. The pangenome consisted of 65 core genes found in greater than 26 strains, 11 soft-core genes present in 25 strains, 252 shell genes present in 4 or more than 4 strains but less than 25 strains, and 161 cloud genes present in less than 4 strains.

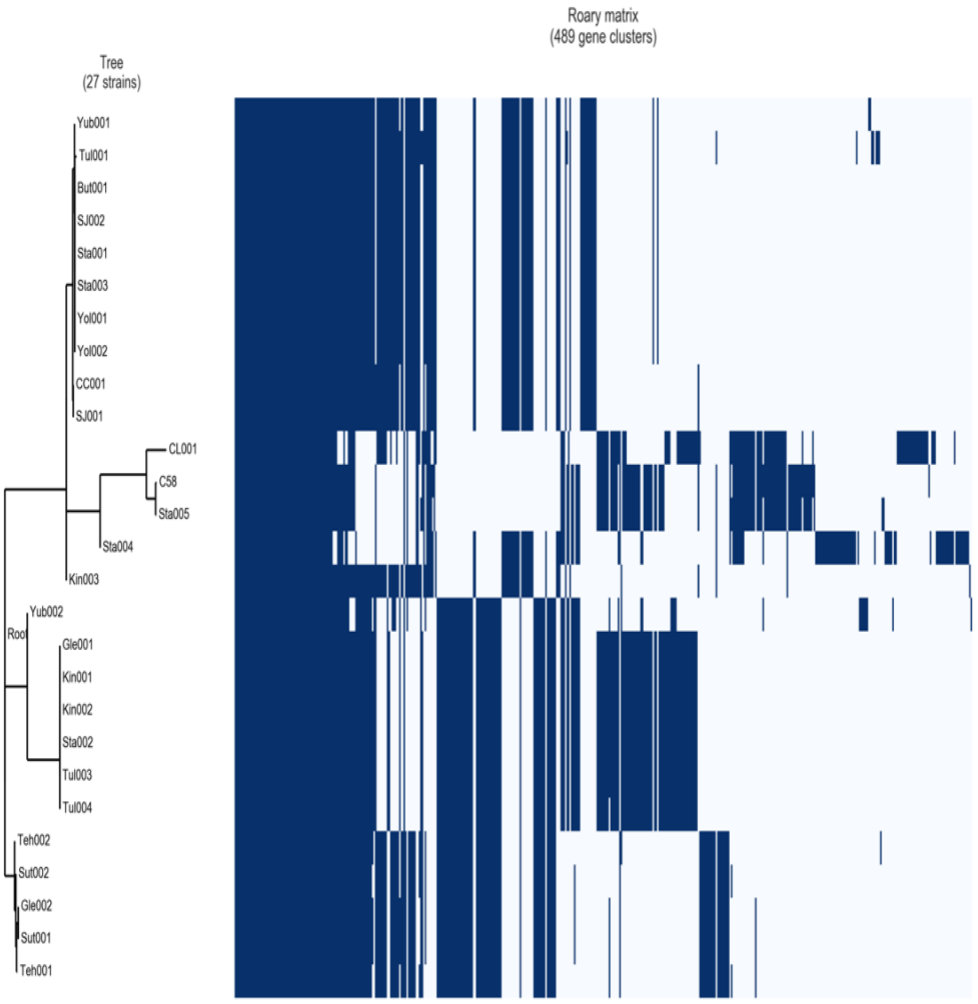


Fig. 4.17. Phylogenetic profile of 27 Ti plasmids with a presence/absence matrix of all genes in the pangenome. The left is the phylogenetic tree of 27 Ti plasmids, aligned by the gene presence/absence profile (right) of each strain. Color blue showed the presence of a gene in a plasmid. Color white represents the absence of a gene in a plasmid.

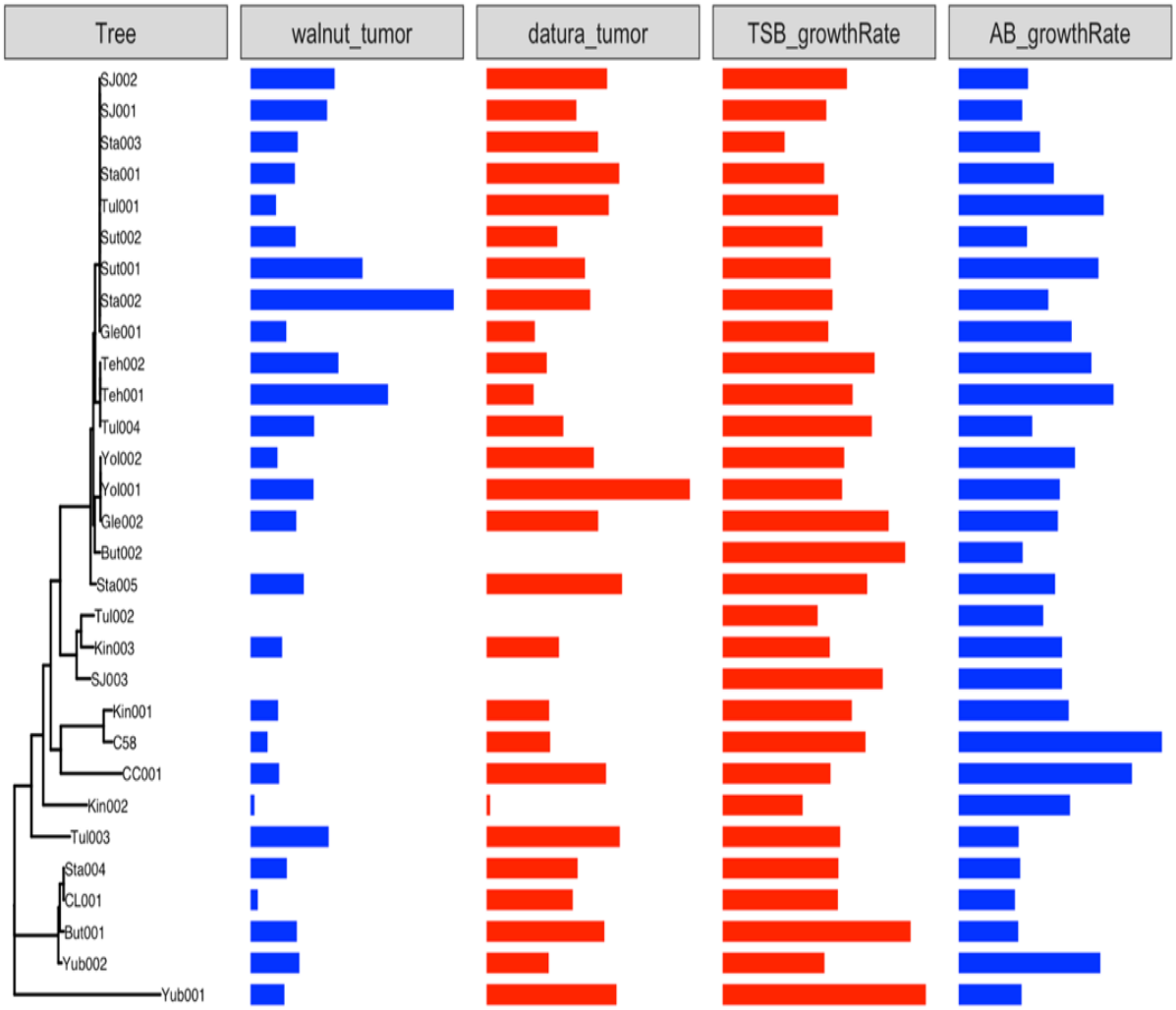


Fig. 4.18. Thirty *A. tumefaciens* strains phylogeny aligned with walnut tumor, datura tumor, growth rate in nutrient rich medium TSB (TSB_growthRate), growth rate in nutrient poor medium (AB_growthRate). The tree bar is the phylogenetic tree rooted at Yub001. The other four bars represent the four phenotypic traits obtained in Chapter 3.

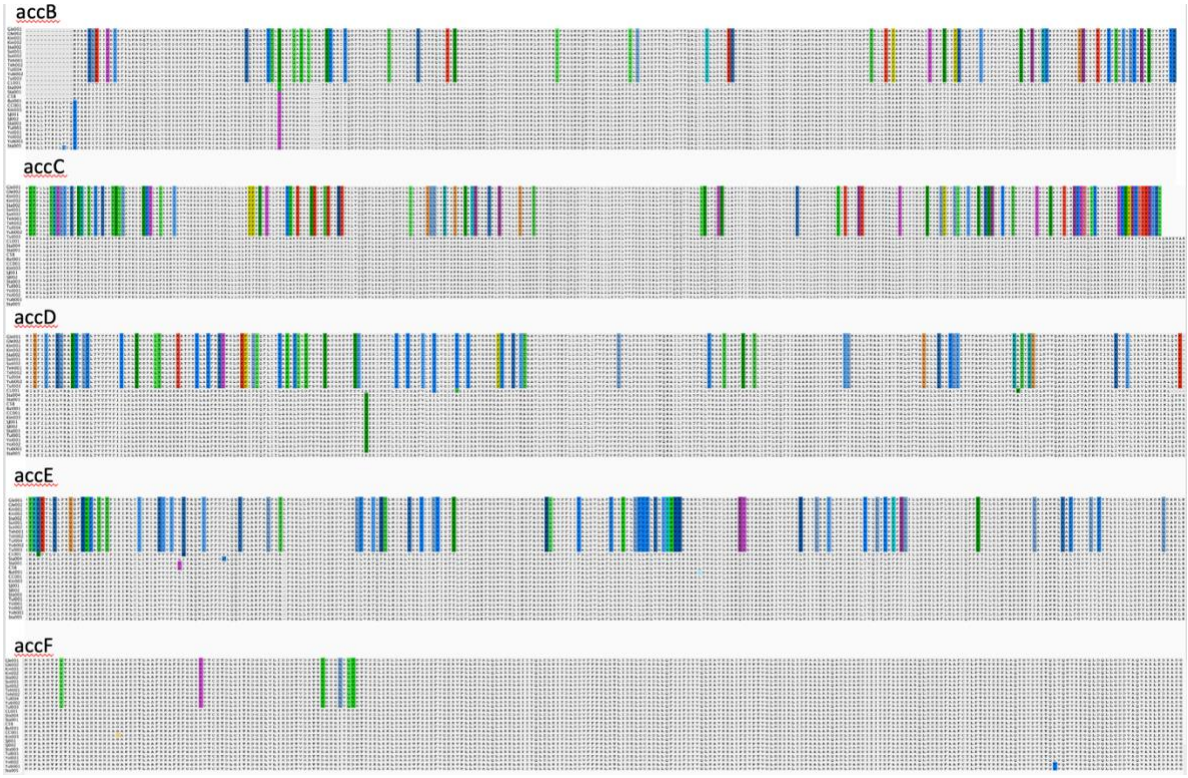


Fig. 4.19. Alignment of amino acids from five genes *accB*, *accC*, *accD*, *accE*, and *accF* of *acc* operon. The 27 virulent *A. tumefaciens* strains tested in Chapter 3 were aligned in the same order (on the left side of each gene alignment). The highlight amino acids represent the amino acids polymorphisms in the genes compared. The strains with highlighted amino acid sequences are resistant to biocontrol strain K84.The rest strains are sensitive to K84.

Note:

Tables for Chapter 4, which contain too much information, are accessible in my github link:

https://github.com/limin321/dissertation_codes/blob/main/chp3%264_codes/Chapter4_tables.xlsx

References

1. Kado, C. I. Historical account on gaining insights on the mechanism of crown gall tumorigenesis induced by *Agrobacterium tumefaciens*. *Front Microbiol* **5**, (2014).
2. Yakabe, L. E., Parker, S. R. & Kluepfel, D. A. Incidence of *Agrobacterium tumefaciens* Biovar 1 in and on 'Paradox' (*Juglans hindsii* × *Juglans regia*) Walnut Seed Collected from Commercial Nurseries. *Plant Disease* **98**, 766–770 (2014).
3. Kim, H. S., Yi, H., Myung, J., Piper, K. R. & Farrand, S. K. Opine-Based *Agrobacterium* Competitiveness: Dual Expression Control of the Agrocinopine Catabolism (acc) Operon by Agrocinopines and Phosphate Levels. *JB* **190**, 3700–3711 (2008).
4. Kerr, A. Biological Control of Crown Gall: Seed Inoculation. *Journal of Applied Bacteriology* **35**, 493–497 (1972).
5. Lacroix, B. & Citovsky, V. Pathways of DNA Transfer to Plants from *Agrobacterium tumefaciens* and Related Bacterial Species. *Annu Rev Phytopathol* **57**, 231–251 (2019).
6. Tempé, J. & Petit, A. 16 - Opine Utilization by *Agrobacterium*. in *Molecular Biology of Plant Tumors* (eds. Kahl, G. & Schell, J. S.) 451–459 (Academic Press, 1982). doi:10.1016/B978-0-12-394380-4.50022-0.
7. Cornforth, D. M. & Foster, K. R. Antibiotics and the art of bacterial war. *PNAS* **112**, 10827–10828 (2015).
8. Ho, B. T., Dong, T. G. & Mekalanos, J. J. A view to a kill: the bacterial type VI secretion system. *Cell Host Microbe* **15**, 9–21 (2014).
9. Cornforth, D. M. & Foster, K. R. Antibiotics and the art of bacterial war. *PNAS* **112**, 10827–10828 (2015).
10. Sterner, R. W. & Elser, J. J. *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere*. (Princeton University Press, 2002).

11. Sundin, G. W. & Wang, N. Antibiotic Resistance in Plant-Pathogenic Bacteria. *Annual Review of Phytopathology* **56**, 161–180 (2018).
12. Pukatzki, S. *et al.* Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. *Proc Natl Acad Sci U S A* **103**, 1528–1533 (2006).
13. Leiman, P. G. *et al.* Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *PNAS* **106**, 4154–4159 (2009).
14. Basler, M., Pilhofer, M., Henderson, G. P., Jensen, G. J. & Mekalanos, J. J. Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**, 182–186 (2012).
15. Wu, H.-Y., Chung, P.-C., Shih, H.-W., Wen, S.-R. & Lai, E.-M. Secretome analysis uncovers an Hcp-family protein secreted via a type VI secretion system in *Agrobacterium tumefaciens*. *J Bacteriol* **190**, 2841–2850 (2008).
16. Lin, J.-S., Ma, L.-S. & Lai, E.-M. Systematic Dissection of the *Agrobacterium* Type VI Secretion System Reveals Machinery and Secreted Components for Subcomplex Formation. *PLOS ONE* **8**, e67647 (2013).
17. VgrG, Tae, Tle, and beyond: the versatile arsenal of Type VI secretion effectors - PubMed. https://pubmed.ncbi.nlm.nih.gov/25042941/.
18. Wu, C.-F. *et al.* Plant-Pathogenic *Agrobacterium tumefaciens* Strains Have Diverse Type VI Effector-Immunity Pairs and Vary in In-Planta Competitiveness. *MPMI* **32**, 961–971 (2019).
19. Bayliss, S. C., Thorpe, H. A., Coyle, N. M., Sheppard, S. K. & Feil, E. J. PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. *GigaScience* **8**, (2019).
20. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution* **26**, 1641–1650 (2009).
21. Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).

22. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research* **44**, 6614–6624 (2016).
23. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
24. Titus Brown, C. & Irber, L. sourmash: a library for MinHash sketching of DNA. *JOSS* **1**, 27 (2016).
25. Shaw, C. H., Watson, M. D., Carter, G. H. & Shaw, C. H. The right hand copy of the nopaline Ti-plasmid 25 bp repeat is required for tumour formation. *Nucleic Acids Research* **12**, 6031–6041 (1984).
26. Wang, K., Herrera-Estrella, L., Van Montagu, M. & Zambryski, P. Right 25 by terminus sequence of the nopaline t-DNA is essential for and determines direction of DNA transfer from *Agrobacterium* to the plant genome. *Cell* **38**, 455–462 (1984).
27. Peralta, E. G. & Ream, L. W. T-DNA border sequences required for crown gall tumorigenesis. *PNAS* **82**, 5112–5116 (1985).
28. Ma, L.-S., Hachani, A., Lin, J.-S., Filloux, A. & Lai, E.-M. *Agrobacterium tumefaciens* Deploys a Superfamily of Type VI Secretion DNase Effectors as Weapons for Interbacterial Competition In Planta. *Cell Host Microbe* **16**, 94–104 (2014).
29. Daly, M., Villa, L., Pezzella, C., Fanning, S. & Carattoli, A. Comparison of multidrug resistance gene regions between two geographically unrelated *Salmonella* serotypes. *Journal of Antimicrobial Chemotherapy* **55**, 558–561 (2005).
30. History. *California Walnuts* https://walnuts.org/about-walnuts/history/.
31. Snapshot.
32. Kluepfel, D. *et al.* DEVELOPMENT OF DISEASE-RESISTANT WALNUT ROOTSTOCKS: INTEGRATION OF CONVENTIONAL AND GENOMIC APPROACHES. 11 (2013).
33. Weisberg, A. J. *et al.* Unexpected conservation and global transmission of agrobacterial virulence plasmids. *Science* **368**, (2020).

34. Vladimirov, I. A., Matveeva, T. V. & Lutova, L. A. Opine biosynthesis and catabolism genes of *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*. *Russ J Genet* **51**, 121–129 (2015).
35. Elser, J. J. *et al.* Biological stoichiometry from genes to ecosystems. *Ecology Letters* **3**, 540–550 (2000).
36. Elser, J. J. *et al.* Growth rate–stoichiometry couplings in diverse biota. *Ecology Letters* **6**, 936–943 (2003).
37. Russell, A. B., Peterson, S. B. & Mougous, J. D. Type VI secretion system effectors: poisons with a purpose. *Nature Reviews Microbiology* **12**, 137–148 (2014).
38. Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. & Attree, I. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics* **10**, 104 (2009).
39. Shi, L., Gao, Z., Zhang, T., Zhang, H. & Dong, Y. Crystal structure of the type VI immunity protein Tdi1 (Atu4351) from *Agrobacterium tumefaciens*. *Acta Crystallogr F Struct Biol Commun* **75**, 153–158 (2019).

Chapter 5

Conclusions and perspectives for future research

This dissertation provides a characterization of the phenotypic and genetic diversity of the *Agrobacterium tumefaciens* complex. Included in this study was an examination of the diversity of strains from the top ten walnut producing counties of the Central Valley of California. In Chapter 2, ANI-based, MLS, core gene phylogenies, and pangenome analyses were used. This analysis resulted in the identification of four clades: clade 1 corresponding to *A. tumefaciens* species, clade 2 being a mix of *Agrobacterium* strains, clade 3 corresponding to *Agrobacterium rhizogenes*, and clade 4 corresponding to *Agrobacterium vitis*. Topology analyses show that whole genome-based phylogenies (ANI-based and core gene) are in agreement in terms of the tree topology, whereas they differed from the 6-gene based MLS phylogeny.

Whole genome ANI analysis of 311 *A. tumefaciens* strains revealed 35 ANI groups when a cut-off of > 95% ANI value between two genomes was used. Each of the above mentioned 4 clades contained more than three ANI groups, indicating substantial genetic diversity in each clade. Comparison of ANI group members with genomic species designation revealed that each ANI group corresponds to a given genomic species for *A. tumefaciens.* For example, genomic species 8 (G8) of *A. tumefaciens*, also known as *A. fabrum* [1], contains the same members as in ANI8. These data support future efforts to assign a formal species name to each ANI group. It would then be appropriate to consider assigning a new genus name to each clade. We propose that this approach of using high quality ANI grouping to define and designate a new species will be applicable to many bacterial genera.

Studies of Chapter 3 and Chapter 4 included 28 strains for California walnuts, one from walnuts in Chile and the type species, C58 for a total of 30 strains. The 29 *A. tumefaciens* strains (28 strains for California walnuts and one from walnuts in Chile) were placed in eight different ANI groups, indicating a high degree of genetic diversity among these *A. tumefaciens* strains. Phylogenetic analysis indicates that strain CL001 from walnuts in Chile was most closely related to the strain Sta004 isolated from a walnut orchard in Stanislaus County in California. Both CL001 and Sta004 were in genomic species 8 (G8). This

observation challenges the idea that genomic species occupies specific ecological niches [2]. The remaining 28 strains were collected from the top ten walnut growing counties of California. Interestingly, genetic diversity was not related to geographic origin of isolation. This is consistent with the walnut nursery trade practice in California where a few major nurseries ship walnut trees to all the major walnut growing regions in California.

The phylogenetic diversity of the strains isolated from California walnut orchards raises questions about the potential phenotypic diversity of these strains and what practical impacts this might have on walnut production. We addressed this question by examining five key phenotypes, i.e., virulence, motility, antibiotic resistance, K84 sensitivity, and growth rates in nutrient-rich and -poor media for the 30 *A. tumefaciens* strains (Chapter 3). We also performed a comparative genomic analysis to investigate the genetic diversity of these strains, the results of which are presented in Chapter 4.

In Chapter 3, the results of examining the five phenotypic traits are presented. Statistically significant differences for these traits were detected. Virulence tests on two hosts, i.e., *Datura stramonium* (*D. stramonium*) and walnut, revealed three avirulent strains and demonstrated significant variation among the 27 virulent strains in terms of tumor sizes. Correlation analysis revealed that the tumor size in walnut and *Datura* are positively correlated. This information may be useful in breeders, when designing a crown gall resistance screening protocols and selecting of appropriate *A. tumefaciens* strains for screening.

In addition to developing disease resistance, selecting effective biocontrol strains to manage *A. tumefaciens* is a worthwhile effort. The observed variation in K84 sensitivity (50% strains were resistant) suggests that use of a single antagonistic strain may not be widely effective. Identification of additional biocontrol strains is needed, along with development of engineering new biocontrol strains. Although the main effective component of a biocontrol strain is its secreted antibiotic, applying conventional antibiotics directly in the field is strictly regulated. Conventional antibiotics are broadly used to control human and

animal disease, and the overuse and misuse of antibiotics contribute to the generation of antibiotic resistant strains. Given that antibiotic resistant genes can be horizontally transferred, application of conventional antibiotics in agriculture should be carefully monitored and regulated, including for crown gall management.

Not only was there variation in virulence, K84 sensitivity and antibiotic resistance among the 30 strains, but there was also significant variation in motility and growth rates. C58 and CC001 grew fastest in the nutrient poor AB medium, with a median doubling time less than two hours and a doubling time of about two hours in TSB. However, correlation analysis of the five phenotypic traits revealed that AB growth rate and carbenicillin resistance, TSB growth rate and erythromycin and rifampin resistance were positively correlated. Growth rate in TSB and walnut tumor size, growth rate in AB and in TSB, and motility and growth rate in AB were negatively correlated (Chapter 3). The variation in phenotypic traits was considered in the genomic comparison of the 30 strains, which is described in Chapter 4.

In Chapter 4, we performed a genomics comparison of the 30 strains (27 virulent and 3 avirulent), focusing on Ti plasmids/T-DNA structure, opine catabolism genes, K84 resistance and T6SS. The Ti plasmid is the primary genetic element responsible for pathogenesis [3]. Genomic comparison of the 27 T-DNA regions of the 27 virulent strains classified them into three opine types: nopaline, agropine, and succinamopine. Association analysis of opine type and tumor size from *D. stramonium* and walnut demonstrate that *A. tumefaciens* strains with different opine types display a degree of plant host specificity. For example, succinamopine strains was more virulent on compared with walnut, whereas agropine strains preferred walnut over Datura.

T-test (alpha = 0.05) results show that opine types have a statistically significant impact on *D. stramonium* tumor size, whereas no significant impact was observed on walnut tumor size. To explain this observation, we hypothesize that the 30 *A. tumefaciens* strains from walnut orchards in California represent different sources and times and also different preferences for 2 host. These host preferences

may explain the significant difference in tumor sizes on *D. stramonium*. After being introduced into walnut orchards, the walnut rootstock becomes their only host available for virulent agrobacteria of different opine types. These strains may have evolved to adapt to this new host, leading to similar size tumors on the walnut rootstock. The t-test (alpha = 0.05) on host genotypes also revealed that host genotypes have a statistically significant impact on tumor size for both agropine and succinamopine opine types. These data provide further evidence that the 30 strains may have host preferences. Therefore, the non-significant difference in walnut tumor sizes may be explained by the idea that Ti plasmids and the host plant are co-evolving to adapt to the walnut rootstock for the succinamopine and agropine type Ti plasmids.

In addition to the T-DNA, the Ti plasmids harbored many opine catabolism genes [4]. Strains with the Ti-plasmid containing agrocinopine catabolism genes are known to be sensitive to the biocontrol strain K84. *Agrobacterium tumefaciens* strains expressing agrocinopine catabolism genes are reported to be sensitive to K84 [4]. The 30 strains we tested all contain genetic loci that are homologous to known agrocinopine catabolism genes on their Ti plasmids. However, half of these strains were resistant to K84. All succinamopine strains were sensitive to K84. Almost all agropine strains were resistant to K84, except strain Tul003. The *acc* operon is one potential mechanism conferring to K84 resistance [5]. Genomic comparison of the *acc* operon from sensitive and resistant strains show different amino acid sequences in accB, accD, accE, and accF that are associated with K84 resistance.

Genomic comparison of the T6SS genes was also performed on 30 strains. Three virulent strains (CC001, Tul003, and Yub001) did not possess a T6SS, indicating that T6SS is not required for pathogenesis. Consistent with this notion, the three avirulent strains did possess the T6SS. The *imp* operon was conserved in 27 T6SSs, whereas the *hcp* operon was highly variable. The 27 T6SSs were classified into 8 groups, which were named through hcp_c1 to hcp_c2 based on the variable gene components of the *hcp* operon. Interestingly, 18 strains contain a second copy of the *tssI* gene, which encodes a protein forming

a spike of the T6SS. Putative T6SS effectors were predicted given their synteny and sequence similarity to known T6SS effectors in other *A. tumefaciens* strains[6] and other systems [7]. This genetic diversity may be important in facilitating *A. tumefaciens* adapting to varied and changing environments.

The above conclusions are derived from bioinformatic data analyses. These *in silico* data analyses suggest many hypotheses for future research to design and perform laboratory experiments or other bioinformatics analyses. For classification of the *Agrobacterium* genus presented in Chapter 2, we propose the concept that ANI group level should be assigned a species name and clade level be assigned a genus name. As more high-quality whole genome sequences become available, we strongly recommend including strains from the Rhizobiaceae family to perform ANI, core genes phylogeny and pangenome analyses to facilitate classification of strains in the *Agrobacterium* genus or Rhizobiaceae family.

My results in Chapter 3 and 4 suggest that *A. tumefaciens* strains have variable T6SS, potential effector-immunity (EI) pairs are predicted. T6SS function and predicted EI pairs need to be examined to understand how they may contribute to the adaptation and survival of *A. tumefaciens*. In addition, the mutation of the *accF* gene of the *acc* operon contributes to resistance to K84 of *A. tumefaciens* strains [5]. Other mutation patterns were found in *accB, accC, accD*, and *accE* genes that may be responsible for the observed resistance to Agrocine 84.

In summary, my dissertation research provides robust information to support reclassification of the genus Agrobacterium based on whole genome sequences. This research also revealed the existence of high level of phenotypic and genetic diversity among pathogenic *A. tumefaciens* strains in California walnut orchards. This information is essential in the continued effort to develop effective and sustainable management strategies for crown gall disease.

References

1. Shams, M., Vial, L., Chapulliot, D., Nesme, X. & Lavire, C. Rapid and accurate species and genomic species identification and exhaustive population diversity assessment of Agrobacterium spp. using recA-based PCR. *Syst. Appl. Microbiol.* **36**, 351–358 (2013).
2. Lassalle, F. *et al.* Genomic species are ecological species as revealed by comparative genomics in Agrobacterium tumefaciens. *Genome Biol. Evol.* **3**, 762–781 (2011).
3. Hooykaas, Paul. J. J. & Beijersbergen, A. G. M. The Virulence System of Agrobacterium Tumefaciens. *Annu. Rev. Phytopathol.* **32**, 157–181 (1994).
4. Vladimirov, I. A., Matveeva, T. V. & Lutova, L. A. Opine biosynthesis and catabolism genes of Agrobacterium tumefaciens and Agrobacterium rhizogenes. *Russ. J. Genet.* **51**, 121–129 (2015).
5. Kim, H. S., Yi, H., Myung, J., Piper, K. R. & Farrand, S. K. Opine-Based Agrobacterium Competitiveness: Dual Expression Control of the Agrocinopine Catabolism (acc) Operon by Agrocinopines and Phosphate Levels. *J. Bacteriol.* **190**, 3700–3711 (2008).
6. Wu, C.-F. *et al.* Plant-Pathogenic Agrobacterium tumefaciens Strains Have Diverse Type VI Effector-Immunity Pairs and Vary in In-Planta Competitiveness. *Mol. Plant-Microbe Interactions®* **32**, 961–971 (2019).
7. Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. & Attree, I. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics* **10**, 104 (2009).