environmental microbiology

Environmental Microbiology (2014) **16**(7), 2112–2125 doi:10.1111/1462-2920.12289

Phylogeography and population structure of the biologically invasive phytopathogen *Erwinia amylovora* inferred using minisatellites

Andreas Bühlmann,[1†] Tanja Dreo,[1,2†] Fabio Rezzonico,[1,3] Joël F. Pothier,[1,3] Theo H. M. Smits,[1,3] Maja Ravnikar,[2] Jürg E. Frey[1*] and Brion Duffy[1,3]

[1]*Plant Protection Division, Agroscope Changins-Wädenswil Research Station ACW, CH-8820, Wädenswil, Switzerland.*

[2]*Department of Biotechnology and Systems Biology, National Institute of Biology (NIB), SI-1000, Ljubljana, Slovenia.*

[3]*Group for Environmental Genomics and Systems Biology, Department of Life Sciences and Facility Management, Institute for Natural Resources Sciences, Zurich University of Applied Sciences, CH-8820, Wädenswil, Switzerland.*

Summary

Erwinia amylovora **causes a major disease of pome fruit trees worldwide, and is regulated as a quarantine organism in many countries. While some diversity of isolates has been observed, molecular epidemiology of this bacterium is hindered by a lack of simple molecular typing techniques with sufficiently high resolution. We report a molecular typing system of *E. amylovora* based on variable number of tandem repeats (VNTR) analysis. Repeats in the *E. amylovora* genome were identified with comparative genomic tools, and VNTR markers were developed and validated. A Multiple-Locus VNTR Analysis (MLVA) was applied to *E. amylovora* isolates from bacterial collections representing global and regional distribution of the pathogen. Based on six repeats, MLVA allowed the distinction of 227 haplotypes among a collection of 833 isolates of worldwide origin. Three geographically separated groups were recognized among global isolates using Bayesian clustering methods. Analysis of regional outbreaks confirmed presence of diverse haplotypes but also high representation of certain haplotypes during outbreaks. MLVA analysis is a practical method for epidemiological studies of *E. amylovora*, identifying previously unresolved population structure within outbreaks. Knowledge of such structure can increase our understanding on how plant diseases emerge and spread over a given geographical region.**

Received 18 July, 2013; accepted 14 September, 2013. *For correspondence. E-mail juerg.frey@agroscope.admin.ch; Tel. (+41) 44783 6424; Fax (+41) 44783 6341. †These authors have contributed equally to the development of this work.

Introduction

Minisatellites, or Variable Number of Tandem Repeats (VNTRs), have been found to be an efficient genotyping method particularly with genetically homogenous species (van Belkum, 1999). Variable number of tandem repeats are tandem repeats of more than five base pairs. Their primary mode of evolution is Slipped Strand Mispairing (SSM) (Torres-Cruz and van der Woude, 2003). The rate of SSM and thus speed of VNTR evolution has been shown to depend on the efficiency of bacterial DNA repair systems (Caporale, 2003). Variable number of tandem repeats have also been proposed to play a role in functional gene evolution by several modes of action. Increasing number of repeats in Shine-Dalgarno sequence can affect gene expression levels (Geluk *et al.*, 1998). Within coding regions, a change in the number of repeats can disrupt the coding frame, resulting in altered length of protein product or change the three-dimensional structure of the gene product as shown in virulence genes of *Haemophilus influenzae* (Hood *et al.*, 1996). These functional effects have led to the assumption that increased genetic variability caused by repeats offer microorganisms evolutionary advantages (Zaleski *et al.*, 2005). Variable number of tandem repeats have been described as fast molecular clocks (van Belkum, 1999) and are thus a perfect tool to type recently emerged bacteria with limited amount of genetic diversity. Dynamics of VNTRs depend on repeat copy number (Vogler *et al.*, 2006) and thus different VNTRs show different clock speeds (Lindstedt, 2005; van Belkum, 2007). Soon after finishing the first bacterial genome of *H. influenzae* (Fleischmann *et al.*, 1995), its sequence information was used to design a VNTR typing system (van Belkum *et al.*, 1997). This application of VNTR-based typing on multiple loci to elucidate population structure has been termed Multiple-Locus

VNTR Analysis (MLVA). The MLVA analysis was subsequently adopted by microbiologists to study population structure of human pathogens such as *Mycobacterium tuberculosis* (Mazars *et al*., 2001), *Yersinia pestis* (Klevytska *et al*., 2001) and *Staphylococcus aureus* (Malachowa *et al*., 2005). Later, genotype information was linked with geographical information to study how bacteria behave within smaller geographical regions or even single outbreaks (Girard *et al*., 2004; U'Ren *et al*., 2007; Bui Thi Ngoc *et al*., 2009; Stratilo and Bader, 2012). Several methods have been implemented to infer mutation rates using stepwise mutation models in *Escherichia coli* (Vogler *et al*., 2006), *M. tuberculosis* (Aandahl *et al*., 2012) and *Y. pestis* (Vogler *et al*., 2007).

Fire blight as a plant disease of pome fruit was first described in New York State in the late 18th century (Denning, 1793) and later, the causative agent, *Erwinia amylovora* was described by Burrill (1882). Its potential for rapid epidemiological spread, and agronomic and economic severity of disease on a wide range of plants of the family of the Rosaceae (especially those belonging to the Spiraeoideae subfamily) have been reviewed (Bonn and van der Zwet, 2000). The lack of commercially effective control strategies other than prophylactic application of antibiotics (i.e. streptomycin) places *E. amylovora* as one of the 10 most important bacterial plant pathogens (Mansfield *et al*., 2012; Stockwell and Duffy, 2012). As a consequence of its recent global spread, genetic diversity of *E. amylovora* is low, ranging from 99.4% identity on the genome level within *Rubus*-infecting strains to 99.98% within the Spiraeoideae-infecting strains (Mann *et al*., 2013). This limited genetic diversity has resulted in difficulties in characterizing different isolates by routine typing methods such as Pulsed Field Gel Electrophoresis (Zhang and Geider, 1997; Jock *et al*., 2002), Polymerase Chain Reaction (PCR) ribotyping (McManus and Jones, 1995), amplified fragment length polymorphism (Rico *et al*., 2004) and clustered regularly interspaced short palindromic repeats (CRISPR) (Rezzonico *et al*., 2011; McGhee and Sundin, 2012). These typing methods can distinguish between isolates from *Rubus* and Spiraeoideae host plants but often fail to distinguish isolates within the Spiraeoideae group, especially within Europe where the diversity is lower (Gehring and Geider, 2012).

An understanding of the evolutionary history and dispersal between and within continents may help to identify inoculum sources in order to prevent invasions of bacterial pathogens such as *E. amylovora* in the future. Information on epidemiology and genetic characterization of individual outbreaks may answer open questions on how bacteria spread artificially on a global scale or are dispersed naturally at the local level, if bacterial populations are substructured, and how to minimize damage during outbreak years (e.g. to protect orchards from neighbouring infected trees). In order to study epidemiology and perform microbial source tracking, it is extremely important to distinguish between genotypes or to assign bacteria to predefined populations. The traditional methods mentioned above did show limited success in achieving the resolution needed for *E. amylovora* and thus, improved molecular typing methods need to be developed (Achtman, 2008). Recent advances in sequencing facilitates identification of molecular markers such as single nucleotide polymorphisms (SNPs), microsatellites and minisatellites (Parkhill and Wren, 2011). In this study, we analyzed the genomic sequences of 12 *E. amylovora* strains to detect VNTR markers with sufficiently high discriminatory power that enable to distinguish isolates within single orchard outbreaks. The following MLVA analysis was performed to gain further insight into: (i) global diversity and population structure of *E. amylovora*, (ii) temporal variation of *E. amylovora* within orchards and individual trees and (iii) potential host association of genotypes and thus adaptation to domesticated hosts. Our results will add to the current understanding on how recently emerged, clonal plant pathogens evolve in different agricultural environments.

Results

Identification of tandem repeat regions and MLVA method development

Using the program JSTRING (De Fonzo *et al*., 2008), a total of 16 tandem repeat regions that met the selection criteria (score = 200, maxgamma = 50, mild penalty) were identified in the genome sequence of strain CFBP 1430. Only one of the 16 tandem repeat regions, which is identical to the previously described Small Repetitive Region (Schnabel and Jones, 1998; Kim and Geider, 1999), was located on the almost ubiquitous plasmid pEA29 (locus A). Thus, an additional repeat on the plasmid pEA29 (locus B), identified during manual sequence assembly of *E. amylovora* CFBP 1430, was included in this study. These two VNTRs on plasmid pEA29 were included in our study since the plasmid is virtually ubiquitous and since the different rate of evolution and different copy numbers between bacterial chromosomes and plasmids might harbour interesting additional information. Since variations in repeat numbers have previously been linked to phenotypic effects giving rise to selective advantages in *H. influenzae* (Hood *et al*., 1996; Geluk *et al*., 1998; Zaleski *et al*., 2005), possible phenotypic effects were evaluated also in the case of *E. amylovora*, however without obtaining any obvious result. Ten of the repeat regions with a repeat unit equal to or larger than 6 bp were tested by PCR on an initial set of 30 strains from diverse geographical origin (Table S1). The highest diversity was

observed in chromosomal loci C and D, while intermediate diversity was observed on chromosomal loci F and H, and few or no differences could be detected at VNTRs E, G, J and K (data not shown). Therefore, the latter four were excluded from further analyses. The remaining six VNTRs are distributed evenly throughout the *E. amylovora* chromosome or the pEA29 plasmid.

Test panel results

Specificity testing on ten related *Erwinia* spp. known to be abundant in the same habitat as *E. amylovora* resulted in only one strain, *Erwinia* sp. 223B from Japan that could be scored on all six loci (Table S2); the resulting haplotype however does not occur in any isolate of *E. amylovora* typed within this study. Thus, it is safe to assume that our VNTR typing scheme is specific to *E. amylovora*. Estimates of individual and pooled diversity index (Hunton–Gaston Discrimination Index, HGDI) values obtained with the six VNTR were determined using the 30 test panel strains. Highest diversity was observed in VNTRs A, C and D, with a diversity index *h* for all strains of 0.79, 0.54 and 0.81 respectively (Table 1). Overall, the high HGDI values obtained for the six VNTRs demonstrate the good discriminatory power of our MLVA design. In the test panel, strain UPN 527 was negative at VNTRs A and B due to the previously reported absence of plasmid pEA29 (Table S2) (Llop *et al*., 2006; Mann *et al*., 2013). Within the test panel dataset, 24 different haplotypes were observed. A minimal spanning tree (MST) of these 30 test panel strains shows higher diversity within North American isolates reflecting the recognized centre of origin of this bacterium (Fig. 1). Additionally, phylogenetic clustering with VNTR markers is in good concordance with CRIPSR clustering (Rezzonico *et al*., 2011) (Fig. 1), indicating absence of horizontal gene transfer and recombination (Feil *et al*., 2001; Spratt *et al*., 2001). The CRISPR group III strains are clearly separated from the group II strains which themselves are outside the group I cluster in the MST tree. While the CRISPR tree fails to resolve group I strains, a better resolution is achieved in the MST tree (Fig. 1). A notable exception in concordance is strain CFBP 3098, isolated in Israel in 1987, which is placed among group I in CRISPR analysis, whereas in our VNTR study, it seems more closely related to group II strains. Bayesian clustering analysis using STRUCTURE shows the highest probability for the existence of two clusters (Fig. 2): (A) a phylogenetically divergent cluster in the United States, including the previously identified CRISPR III group strains (Rezzonico *et al*., 2011) and (B) a cluster comprising phylogenetically closely related strains from the United States and Europe.

VNTR characteristics on a global scale

On a global scale, a total of 227 haplotypes were detected among 833 tested isolates. All loci were polymorphic with allele numbers N_a per locus ranging from minimum 4 for locus B to maximum 16 for locus D (Table 1). Few isolates lacking the plasmid pEA29 did not produce amplicons for VNTR A and B. Two isolates from Spain (IVIA 1614, UPN 527), previously described as devoid of pEA29 (Llop

Table 1. Characteristics of VNTRs from all isolates, the Slovenian subset and the test panel subset. Ni: number of isolates; Na: number of different alleles; Ne: number of effective alleles; I: Shannon's Information Index; h: diversity index (all calculated in GenAlEx). VNTR A and B are on plasmid pEA29, VNTRs C,D,F,H are on the chromosome.

Locus	CFBP 1430 start position	Locus tag	Target region	Subset	N_i	N_a	*Ne*	*I*	Number of repeats Min	Max	*h*
VNTR A	26041	Intergenic	repeat region on pEA29[a]	All	832	13	6.121	1.975	2	13	0.84
				Slovenia	507	9	5.199	1.761	3	11	0.81
				Test panel	29	8	4.861	1.759	4	12	0.79
VNTR B	14703	Intergenic	repeat region on pEA29	All	830	8	1.818	0.709	1	13	0.45
				Slovenia	504	6	1.11	0.257	1	13	0.1
				Test panel	29	2	1.998	0.693	2	3	0.5
VNTR C	457256	Eamy_0389	hypothetical protein	All	830	9	2.705	1.248	3	11	0.63
				Slovenia	508	6	1.493	0.712	5	10	0.33
				Test panel	30	5	2.163	1.036	6	11	0.54
VNTR D	1254089	Eamy_1186	hypothetical protein	All	834	13	3.141	1.494	1	18	0.68
				Slovenia	508	9	1.762	0.899	1	18	0.43
				Test panel	29	8	5.224	1.807	6	16	0.81
VNTR F	2668446	Eamy_2580	*yfgA*	All	834	7	1.307	0.551	4	10	0.23
				Slovenia	508	6	1.205	0.423	4	9	0.17
				Test panel	29	2	1.488	0.51	8	9	0.33
VNTR H	3517853	Eamy_3423	hypothetical protein	All	838	9	1.219	0.463	3	11	0.18
				Slovenia	508	8	1.147	0.339	3	11	0.11
				Test panel	30	3	1.59	0.639	5	8	0.37

a. Repeat region spans the repeat region of SSR with a repeat of GAATTACA motif first described by Schnabel and Jones, 1998.

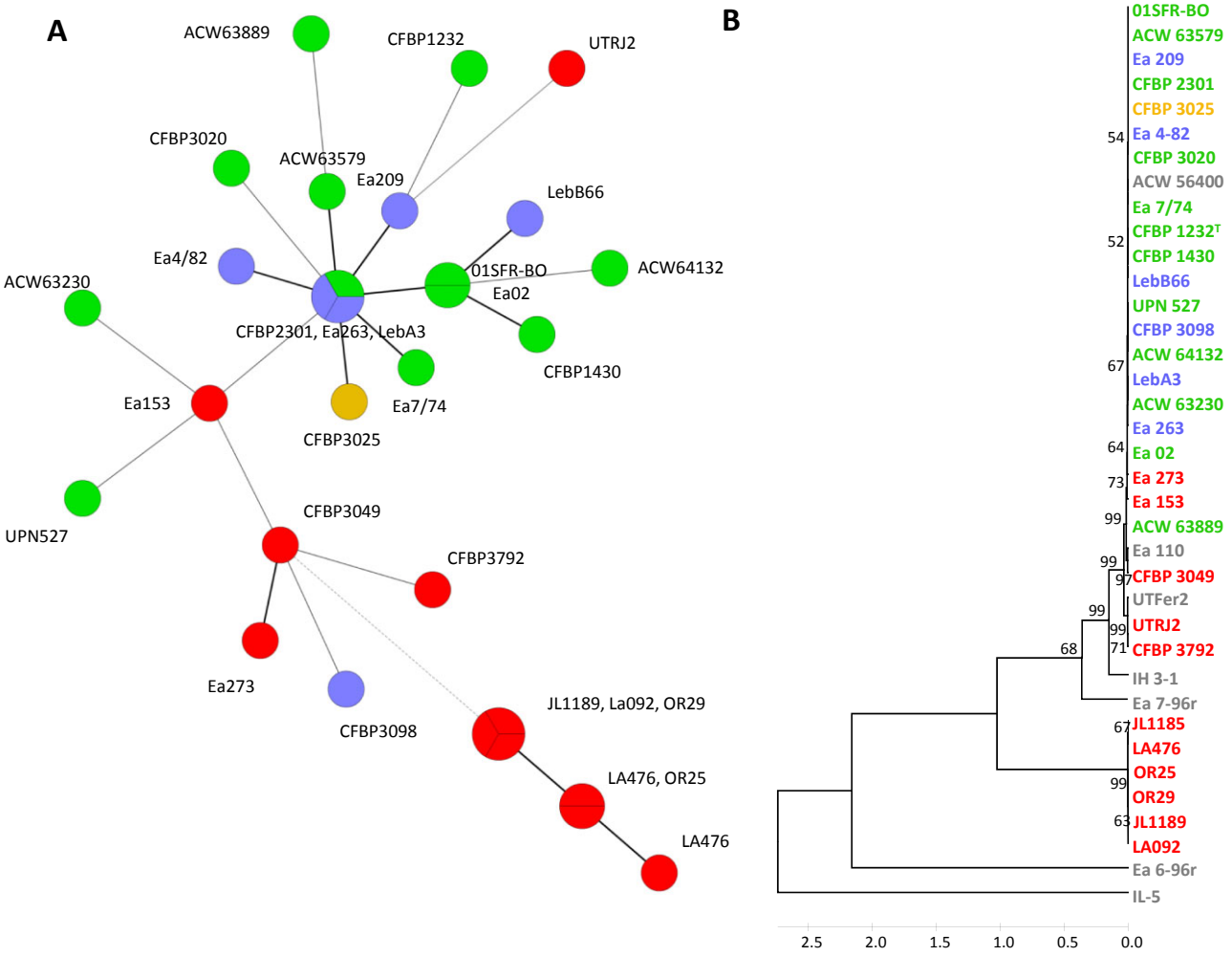

Fig. 1. Comparison of Minimal spanning tree of 30 test panel strains (A) with CRISPRs spacers tree (B) (Rezzonico *et al.*, 2011) of the same set of isolates. Green indicates strains from Europe, red from the USA and blue from the Middle East. Grey indicate strains used in the CRISPR study (Rezzonico *et al.*, 2011), which were not used in the present study.

et al., 2006), two from the United States (AFRS 258, LA469) and one from Slovenia (NIB 1117), confirming that plasmid pEA29 is almost ubiquitous within *E. amylovora* (Table S1). Some isolates from *Rubus* hosts failed to amplify VNTR C due to genetic diversity at this locus (Mann *et al.*, 2013). As in the test panel, the highest diversity was observed in VNTRs A, C and D, with diversity indices for all isolates of 0.84, 0.63 and 0.68 respectively. Minimal spanning trees show a high diversity in US isolates compared to European isolates, and also subdivision into two major clusters within Europe (Fig. 3). Bayesian clustering was performed in STRUCTURE supporting these three groups on the global scale (Fig. 2) that could not be resolved using only the strains in the test

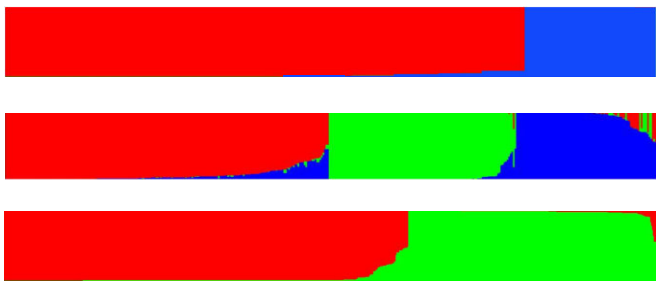

Fig. 2. STRUCTURE outputs for a test panel of 30 strains, all 833 isolates and 733 European isolates. Best K, the true value for number of clusters, was selected using the Evanno method (Evanno *et al.*, 2005). Colours represent groups identifiable by Bayesian clustering.

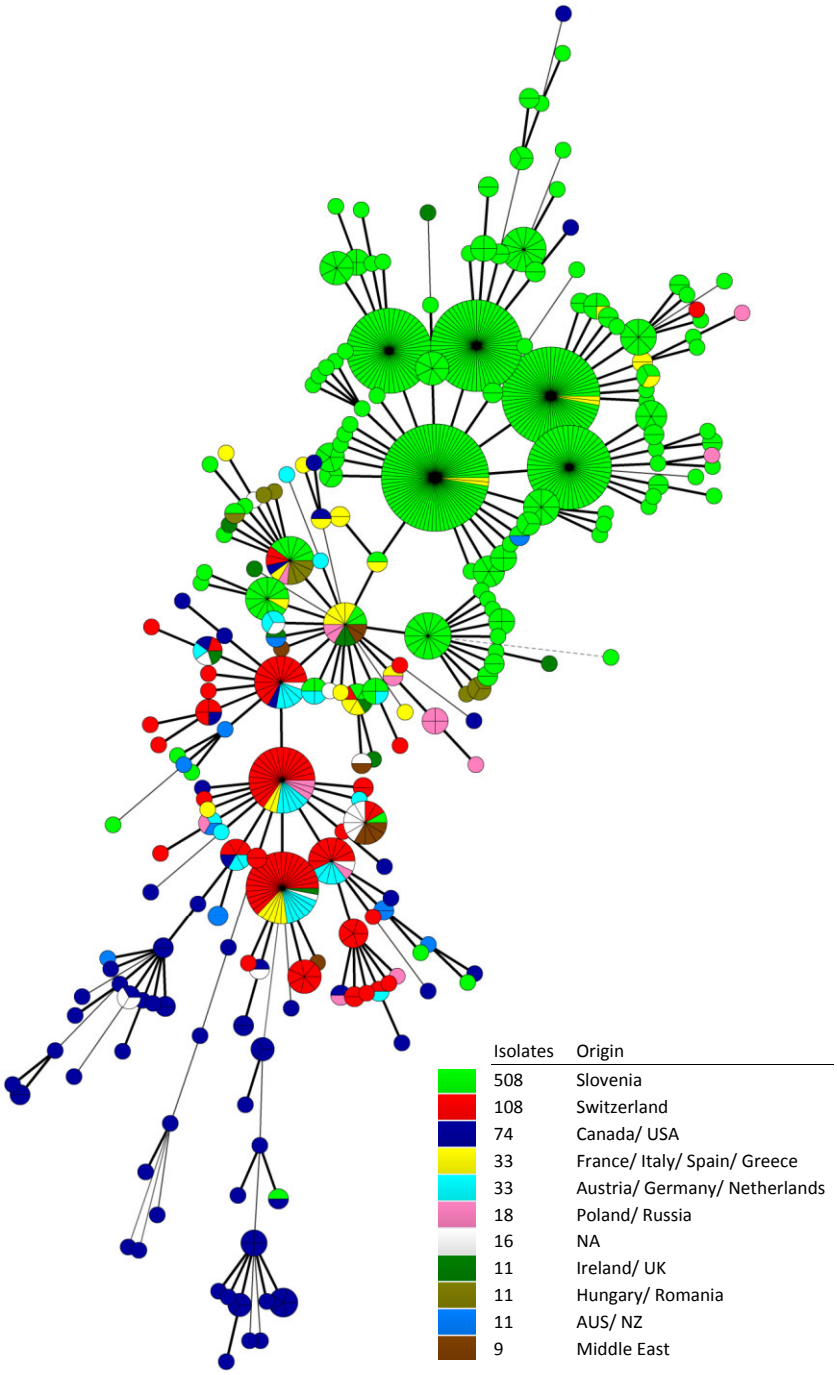

Fig. 3. Minimal spanning tree under a stepwise mutation model on all isolates showing diversity within US isolates (blue) and subdivision into two groups within Europe (green, Slovenia) and (red, Switzerland). Solid bold line: one difference in VNTR repeat number; dashed line: two differences; solid thin line: three differences.

panel above. Analysis of these groups reveals a distinction between a population with isolates covering mostly Northwestern Europe and a population covering Southeastern Europe, mostly Slovenia, Italy and Spain, although populations overlap to some extent showing admixture. Pairwise comparison of Nei's genetic identity confirms a clear differentiation of the three groups (Table 2). The southeastern European group shows different allele frequencies from the rest of European iso-

Table 2. Pairwise population matrix of Nei's genetic identity calculated in GenAlEx, between the three clusters identified by Bayesian clustering in STRUCTURE, showing distinguishable populations. Ni: number of isolates.

	Ni	Group 1	Group 2	Group 3
Group 1	180	1.00		
Group 2	138	0.77	1.00	
Group 3	439	0.65	0.40	1.00

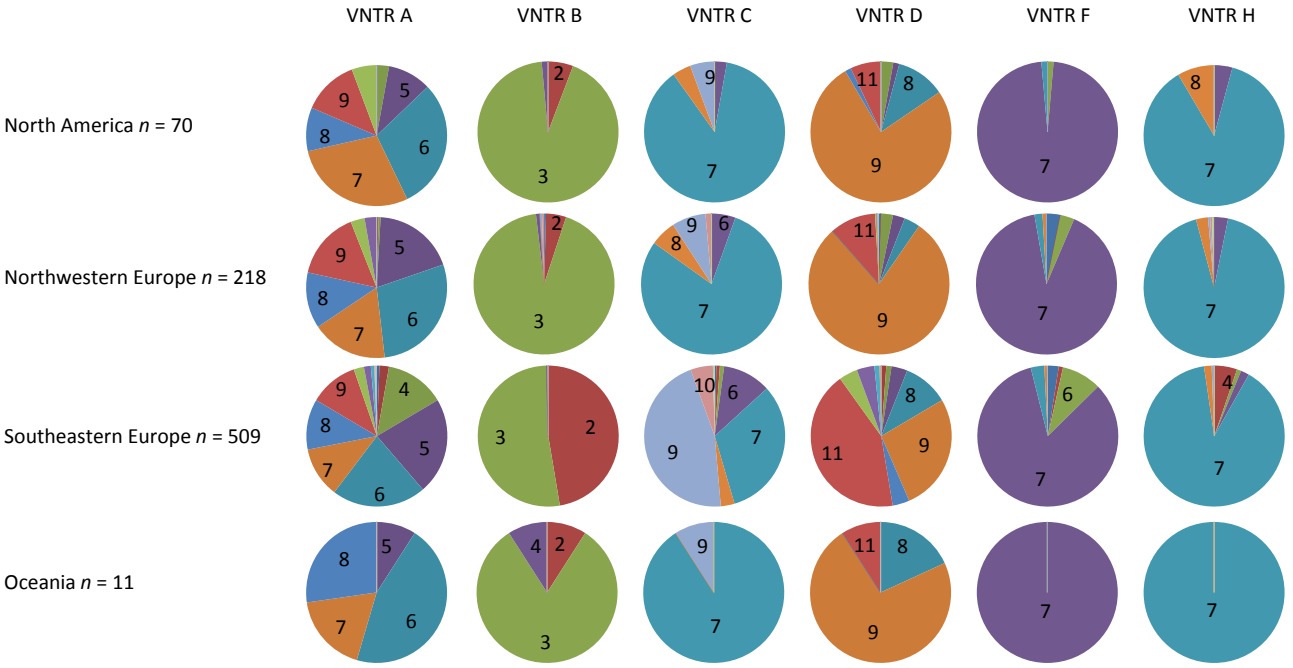

Fig. 4. Allele frequencies of populations from different origin for each locus. Numbers delineate VNTR allele. While VNTRs A, F and H show very similar allele frequencies among populations, VNTRs B, C and D show a clear differentiation between North America, northwestern Europe and southeastern Europe.

lates in VNTRs B, C and D (Fig. 4). Isolates from the Middle East countries Egypt, Israel and Lebanon cluster with the northwestern European group suggesting that *E. amylovora* was introduced into these countries during the same period as into the northwestern European countries or was imported from northwestern Europe to the Middle East. Isolates from New Zealand show a rather diverse pattern, clustering with strains from North America and Europe alike, supporting reports about earlier introduction events of *E. amylovora* to New Zealand.

Analysis on a national scale: Slovenia and Switzerland

Among the 407 Slovenian isolates tested from years 2002 to 2009, most isolates represented extensive outbreaks in 2003 (162 isolates) and 2007 (168 isolates). To minimize sampling bias, only one isolate was sampled per orchard per year. In total, 69 haplotypes were identified with five of them representing 71% of all isolates. One of the most common haplotypes also detected in the first positive sample in 2002 was the predominant strain of both extensive outbreaks and most other years. Compared to the global collection of isolates, lower diversity was observed (Table 1), confirming that only a subgroup of haplotypes is present in Slovenia. During the extensive outbreaks, higher diversity at all loci and a higher proportion of private haplotypes was observed. However, allele frequencies indicated no apparent trend to specific haplotypes between years (Fig. S1), analysis of molecular variance (AMOVA) with permutation tests showed no significant difference of Slovenian populations between years (Table 3), and a minimal spanning tree based on VNTR patterns of Slovenian isolates showed no obvious pattern of year to year differences (Fig. 5). These results indicate that the same strains infect and cause outbreaks every year. A Mantel test of total genetic distance and geographic distance showed weak correlation for Slovenia and Switzerland, implicating weak isolation by distance within a single country, compared to the global level where isolation by distance can be identified (Table 4). Additionally, Bayesian clustering on the same samples failed to identify multiple groups indicating no

Table 3. AMOVA results of Slovenian isolates (Ni = 476)[a] grouped by year showing no significant population differentiation by year.

Source	Est. var.	%	φ_{PT}	*P*(rand ≥ data)[b]
Among populations	0.03	3		
Within populations	0.96	97		
Total	0.99	100	–0.01	0.47

a. Slovenian sample set consisted of, 2002: 4 isolates, 2003:198 isolates, 2004: 31 isolates, 2005: 8 isolates, 2006: 4 isolates, 2007: 192 isolates, 2008: 33 isolates, 2009: 6 isolates.
b. *P*(rand ≥data), Probability of φ_{PT} based on 999 standard permutations across the full dataset.
Est. var.: estimated variance; φPT, measure of population differentiation.

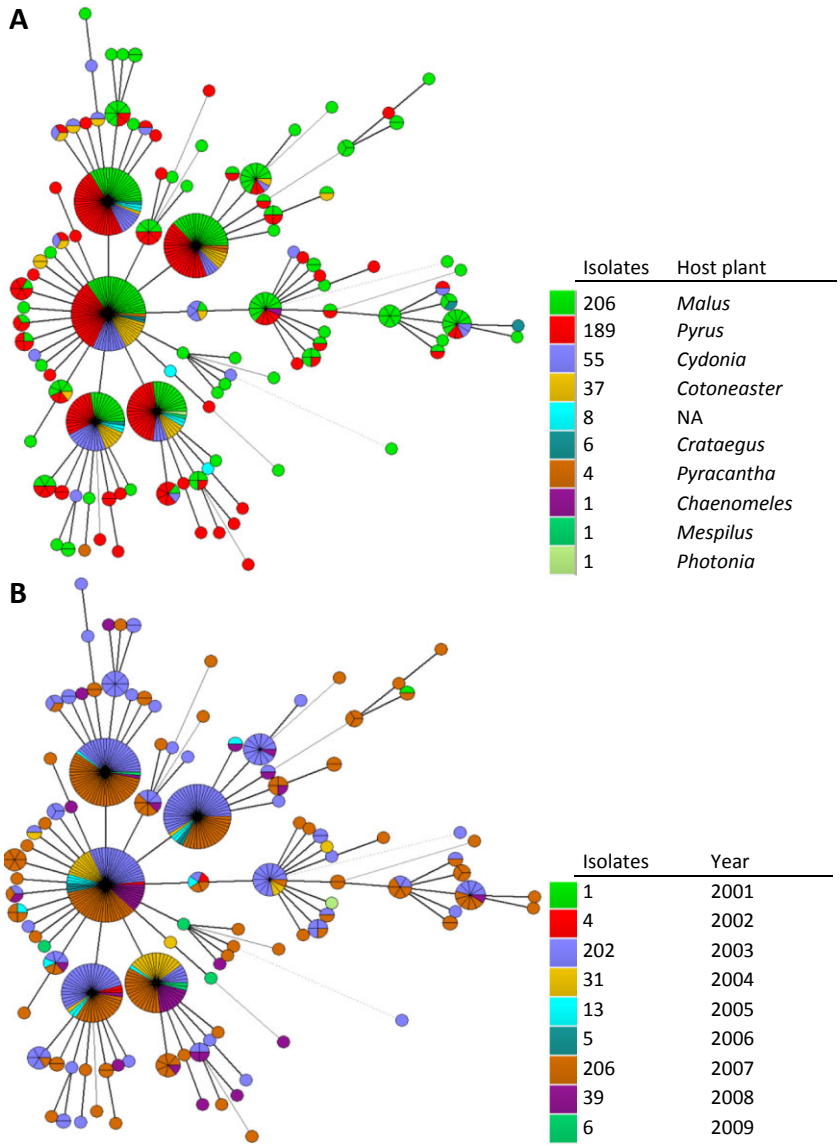

Fig. 5. Minimal spanning tree of 497 isolates grouped by host plants (A) and year of isolation (B). No obvious trend is visible as confirmed by AMOVA and Mantel tests indicating no population subdivision by outbreak year or by association with host plant.

geographical barriers or not enough time to form separate groups. When plotting haplotypes according to sampling location on a map, no trend was apparent confirming these findings (Fig. S2).

Analysis of individual trees

Analysis of 35 isolates of a single naturally infected tree sampled repeatedly over 8 weeks, revealed multiple

Table 4. Mantel test results of various groups. On a global scale, genetic and geographic distances are correlated due to infrequent transfer of strains across continents, whereas in Slovenia and Switzerland, there is less correlation.

Population	SSx	SSy	SPxy	Rxy	p (rxy-rand ≥ rxy-data) [a]
Global	35418110.38	176441.77	1087073.50	0.44	0.001
Slovenia	10518036.96	19625.95	25436.18	0.06	0.008
Switzerland	36001.76	643.67	881.76	0.18	0.001

a. *P* (rxy-rand ≥ rxy-data), probability of Rxy based on 999 standard permutations across the full dataset.
SSx, sum of products of *x* matrix elements; SSy, sum of products of y matrix elements; SPxy, sum of cross products of corresponding elements of the x and y matrices; Rxy, Mantel correlation coefficient.

VNTR types demonstrating feasibility of our VNTR typing scheme to distinguish single orchard outbreaks (Table S3) without extensive DNA isolation procedures. In five of seven sampling weeks, we detected multiple VNTR types suggesting single trees are typically infected by strains belonging to different haplotypes. This result also proves the feasibility of MLVA for direct field sample testing without preceding culture methods.

Host association

Host association and selection towards hosts has been described in several plant pathogens such as xanthomonads (Bogdanove *et al*., 2011) and pseudomonads (Baltrus *et al*., 2011; Gironde and Manceau, 2012). Host range is an important parameter in emergence of bacterial phytopathogens and thus understanding host range can generate valuable information for control and management. To test if host plants drive divergence of *E. amylovora* populations, the global dataset was reduced to a single country with short history of *E. amylovora* infection to exclude changes in allele frequencies by genetic drift or selection. We compared 497 Slovenian isolates from various host species for population subdivision. A minimal spanning tree showed no apparent pattern (Fig. 5). Analyses of molecular variance indicated 100% of total genetic variation within host and thus no evidence for population subdivision. Bayesian clustering did not split isolates from different hosts into different groups. These results imply that *E. amylovora* is equally well adapted to these hosts, and thus the bacteria can spread freely among them.

Discussion

In this study, we identified VNTR loci in *E. amylovora*, evaluated the utility of MLVA as a superior genotyping method for *E. amylovora* compared to MLST and CRISPR analysis (Rezzonico *et al*., 2011; McGhee and Sundin, 2012), and assessed its usefulness for future epidemiological investigations. We used 12 sequenced *E. amylovora* genomes (Smits *et al*., 2010; Mann *et al*., 2013) to develop six VNTR markers able to discriminate isolates from globally diverse geographical origins but also from single tree isolates. The results indicate that all VNTR loci are sufficiently stable to be used in epidemiological tracking studies. Although common VNTR types occur, the fact that the number of alleles at different VNTR loci ranged from 7 to 13 suggests that theoretically $> 500\,000$ haplotypes may be distinguished from each other, resulting in huge resolution of this typing method compared to previously used ones. Multiplexing renders the method comparably cheap since only one multiplex PCR has to be performed per isolate and PCR can be run directly on bacterial samples without extensive DNA purification.

We confirm previous reports that *E. amylovora* genetic diversity is much higher in North America and thus indicating its centre of origin (Rezzonico *et al*., 2011), although overall genetic diversity remains still very low compared to other pathogens. A genetic bottleneck associated with spread of only few genotypes outside North America resulted in dramatically reduced diversity in Europe, New Zealand and the Middle East (Rezzonico *et al*., 2011). The limited amount of time since global spread of *E. amylovora* was not enough to acquire many mutations (Mann *et al*., 2012; 2013). A higher diversity within Europe could be a sign for selection, or for an increase in the molecular clock rate, which is not expected considering the presence of the similar host plant species and even cultivars for commercially important plants. For European isolates, STRUCTURE analysis groups isolates of *E. amylovora* into two major clusters, one cluster covering northwestern Europe, Russia and the Middle East and a southeastern Europe population covering Slovenia, Italy and Spain. Since we failed to identify isolation by distance within Europe, we hypothesize multiple introduction events causing this split population. These population differences were previously unidentifiable in similar studies, although some population structure has been elucidated before (Jock *et al*., 2002; 2013; Donat *et al*., 2007). While the MLVA analysis showed a strong evidence for two populations within Europe, this finding has yet to be confirmed by other methods such as high-throughput SNP typing or next generation sequencing. If so, further experiments could give insight into modes of population subdivision within plant pathogenic bacteria. The linkage disequilibrium calculations showed non-random association between loci, which is expected in clonal bacteria (Smith *et al*., 1993; Supply *et al*., 2003; Spratt, 2004). Some but not widespread Horizontal Gene Transfer (HGT) has been reported for *E. amylovora* (Mann *et al*., 2013), whereas other plant pathogenic bacteria show much higher frequency of HGT reflecting different life cycles (Bogdanove *et al*., 2011; Gironde and Manceau, 2012).

Although MLVA for phylogenetic analysis might not be ideal over very long time periods due to the variable nature of VNTRs, the clustering analysis on genetically diverse isolates showed good concordance to previously generated phylogenies based on CRISPR spacers (Rezzonico *et al*., 2011). Additionally, increased resolution of MLVA allows for deeper phylogenies than previously possible. This higher resolution is critical to determine dispersal of bacterial pathogens within single countries or even within single orchards given that traditional methods failed to distinguish between closely related isolates.

However, CRIPSR and especially MLST analyses remain useful because more information on the evolutionary (MLST) and ecological (CRISPR) trajectory can be generated with these typing schemes. Only limited MLST studies have been performed in *E. amylovora* (Rezzonico *et al*., 2012; Mann *et al*., 2013). Due to the genetically monomorphic nature of this bacterium, mainly differences in *Rubus* – Spireoidae strains could be distinguished in these studies, a distinction that the presented MLVA can achieve as well, with additional resolution within Spireoidae strains. It can be foreseen, that sequencing of whole genomes will soon enable better comparisons of different typing methods (Medini *et al*., 2008).

The MLVA analysis showed no prevalence of a certain genotype in a particular host plant indicating that isolates of *E. amylovora* from Spiraeoideae are generally capable to affect other members of the Spiraeoideae family. Alternative hosts such as *Cotoneaster* or *Pyracantha* spp., which form a genetically more diverse group of host plants with respect to commercial apples, were affected by the same MLVA genotypes found in orchards. This may suggest that the mechanisms of pathogenicity of the Spiraeoideae-infecting strains of *E. amylovora* are suited for a broad range of host plants in contrast to other plant pathogens such as xanthomonads and pseudomonads (Bogdanove *et al*., 2011; Gironde and Manceau, 2012).

We failed to detect significant diversity of haplotypes within outbreaks between years suggesting that no single but multiple introductions of *E. amylovora* occur in orchard ecosystems. Once introduced, several haplotypes remain active within geographic regions and either yet undiscovered mechanisms or chance decide which of the underlying genotypes becomes dominant during an outbreak. This finding reflects the previously described idea that *E. amylovora* overwinters within the host plant, from which it can emerge the following year by spreading of bacterial ooze by insects (Thomson, 2000). Additionally, we did not detect any correlation of genetic and geographic distance confirming the known concepts of *E.amylovora* epidemiology that the bacteria can be spread over relatively long distances by insects (Thomson, 2000). However, our results also highlight a disadvantage of neutral markers in characterising outbreaks. Novel mutations giving rise to selective advantages will by definition be missed in a typing system based on neutral markers. A further shortcoming of studies in bacterial plant pathology can be sampling bias of isolates. To reduce the possibility of sampling bias, only one strain was sampled per orchard in a given year. It still remains to be answered though if and to what extent *E. amylovora* is present on other host plants from environments of apple production areas, on which the microorganism may or may not exhibit larger genetic diversity. It is, however, difficult to sample such environments in an economically reasonable way.

It can be stated that, no matter which genotyping method is used, *E. amylovora* remains a highly monomorphic pathogen at the genomic level (Mann *et al*., 2013). This is probably due to its relatively young evolutionary age and to recent bottleneck events (Rezzonico *et al*., 2011; 2012), whereas the role of the absence of selective pressures remains speculative. Other pathogens of fruit trees, such as *Venturia inaequalis,* a fungal pathogen expected to have coevolved during domestication of apple and pear, were shown to have a panmictic population structure on domestic apple (Gladieux *et al*., 2010). Thus, together with findings on *E. amylovora,* such studies can serve as a working hypothesis for absence of induced selection by genetically monomorphic host plants. However, the absence of genetic diversity raises questions on how *E. amylovora* managed to be such an effective pathogen on a broad range of host plants. Differences at the expression level of virulence genes between strains have been reported (Wang *et al*., 2010), although the genetic background of these differences remains to be elucidated.

The fact that multiple genotypes of *E. amylovora* exist in all regions with established fire blight disease, suggests that spread of similar bacterial pathogens can only be prevented by intensive controls of commercially traded plant material. Lack of effective control strategies heightens the need for effective monitoring and genetic typing systems as described in this study. We demonstrate the efficacy and efficiency of our VNTR typing strategy and the power of integrative genomics to rapidly identify markers and develop genotyping systems for global, regional and on-site diversity analysis as reviewed before (Davey *et al*., 2011). While increasing affordability of direct genomic sequencing will facilitate more profound population genetic study and microbial forensics (Harris *et al*., 2010; Croucher *et al*., 2011; Mutreja *et al*., 2011), VNTR molecular marker typing currently offers the most efficient option for processing large sample sizes required in epidemiological and source tracking investigation of *E. amylovora*.

Experimental procedures

Bacterial isolates

A total of 833 bacterial isolates of *E. amylovora* from fire blight symptomatic plant material of different host plants were used in this study. Isolates originate from public and private bacterial collections. Collections from Slovenia and Switzerland were obtained in the frame of official fire blight monitoring from 2001–2009. All isolates from Slovenia were identified by colony morphology (Luria Bertani and King's B medium), serum agglutination (EamC antibodies, Prime Diagnostics antibodies, Wageningen, The Netherlands),

real-time PCR (*amsC* target, Pirc *et al*., 2009) and pathogenicity test in immature pear fruits. The Swiss isolates were selected on the basis of colony morphology (Levan and King's B medium) and the *E. amylovora* Agristrip immunological kit (Bioreba AG, Rheinach, Switzerland) (Braun-Kiewnick *et al*., 2011). Isolates from other countries were identified according to internal procedures of the originating laboratories; additionally the identity of the isolates was confirmed by real-time PCR targeting *amsC* gene (Pirc *et al*., 2009). Each isolate represents an individual plant sampled either on different orchards or on the same orchard but in a different year. The only exceptions are the 38 single tree replicates to test diversity on single plants. The test panel used in the development of the MLVA scheme consisted of 30 strains of different geographical origin covering the geographic and genetic diversity of *E. amylovora*. A set of other bacterial species that commonly occur on fire blight host plants were included in the analysis to confirm the specificity of the MLVA assay (Table S2).

In-silico *VNTR prediction*

The genome sequence of CFBP 1430 (Smits *et al*., 2010) was used as the reference strain for the bioinformatic analyses using the tandem repeat finder program JSTRING (De Fonzo *et al*., 2008), and VNTRs were confirmed by comparing to the publicly available genome sequence Ea273 in the Tandem Repeats Database (Denoeud and Vergnaud, 2004).

DNA extraction

For MLVA analysis of isolates, single bacterial colonies were subcultured on King's B medium at 25–28°C for 1–2 days. Suspensions of bacteria were prepared in sterile water and were used for DNA extraction using Chelex 100 (Bio-Rad, Hercules, CA, USA) (Walsh *et al*., 1991). Briefly, 1 ml of bacterial suspension was prepared in water with OD_{650} of 0.1–0.2. After vortexing, bacteria were pelleted with 5 min centrifugation at 9900 *g*. The pellet was then mixed with 300 µl of Chelex water suspension (6% w/V), vortexed briefly, incubated for 20 min at 56°C, briefly vortexed, incubated for 8 min at 99°C, vortexed for 10 s, cooled immediately on ice and centrifuged for 7 min at 14 000 *g*. The supernatant-containing DNA was collected (200 µl) and DNA concentrations estimated by Nanodrop (ThermoScientific, Waltham, MA, USA). The Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) was used to extract DNA from symptomatic plant material.

DNA sequencing of VNTR loci

Prior to sequencing, PCR products (3–5 µl) were resolved on agarose gels. Remaining PCR products (usually 5 µl) were cleaned using MultiScreen PCR µ96 plates (Merck Millipore, Billerica, MA, USA) and reconstituted in 20 µl of water. Sequencing reactions contained 2 µl BigDye terminator v1.1 cycle sequencing kit (ABI, Foster City, CA, USA), 0.2 µl of primer (10 µM working concentration) and 5.8 µl of purified PCR product. Sequencing was performed on ABI Prism 3130xl Genetic Analyzer. Sequences were analyzed using SeqMan Pro v.10.1 software (DNASTAR, Madison, WI, USA).

Design and development of MLVA scheme

Ten primer pairs targeting single-locus VNTRs were designed from the full genome sequence of *E. amylovora* CFBP 1430 (Smits *et al*., 2010). The specificity of these primer pairs was evaluated by screening 12 sequenced genomes of *E. amylovora* and by searching the BLAST (nr/nt) database. These non-labelled primer pairs were used in preliminary analyses using single-plex PCR followed by gel electrophoresis and sequencing on a selection of eight strains of different geographic origin (CFBP 1430, CFBP 1332^T, ACW 56400, 01SFR-BO, UPN 527, Ea273, Ea 263 and OR 29). Further analyses on selected strains were done to confirm the diversity of individual VNTRs and their suitability for epidemiological studies. Based on these results, six primer pairs were selected for development of MLVA using capillary electrophoresis, two targeting plasmid pEA29 loci and four targeting loci on chromosomal DNA. One of each primer in each primer pair was labelled with one of the fluorescent dyes FAM, NED, PET, and VIC (Table S4). Polymerase chain reaction products of loci with same dye were adapted to maximize fragment size differences. All MLVA multiplex PCR were performed by mixing 25–60 ng of genomic DNA with 5 µL of Multiplex Master Mix (Qiagen, Hilden, Germany) and 1 µL of primer mixture. The primer mixture was optimized for amplification of all loci and contained 0.35 µM VNTR-A, 0.10 µM VNTR-B, 0.60 µM VNTR-C, 0.35 µM VNTR-D and VNTR-H, and 0.30 µM VNTR-F primers. Polymerase chain reaction amplifications were performed on a GeneAmp PCR system 9700 thermocycler (Applied Biosystems, Foster City, CA, USA) under the following conditions: 15 min at 95°C for hot-start activation; 30 cycles of 94°C for 30 s, 56°C for 90 s, 72°C for 4 min; and a final extension step at 72°C for 30 min. To ensure reproducibility strain, CFBP 1430 was run in every PCR plate as positive control.

Capillary electrophoresis

The PCR products of multiplex MLVA were cleaned using MultiScreen PCR µ96 plates (Merck Millipore) and DNA reconstituted in purified water. Aliquots of 1 µl of cleaned products (DNA) were diluted 1/100 and mixed with 8.8 µl of Hi-Di™ formamide (Applied Biosystems) and 0.07 µl of a GeneScan 500 LIZ internal lane size standard (Applied Biosystems). Standard capillary electrophoresis was performed in an ABI PRISM 3130xl genetic analyser (Applied Biosystems). Results were scored and analysed using GeneMapper v 4.0 and PeakScanner v1.0.

Data scoring and description of MLVA profiles

The length and number of repeats was determined in 30 isolates from the test panel (Table 1) through sequencing. The number of repeated motifs was rounded to the next highest integer number. As such, the null allele designates a locus that contains both flanking sequences but no repeat unit. If more than one amplicon for a specific VNTR locus was

detected, and the size difference matched more than the repeat lengths (so-called stutter peaks), the one with the highest fluorescence level was used to assign the repeat number. An allele number string, based on the concatenated repeat numbers, was assigned to all isolates describing unambiguous identification of allele types (Table S5). After fragment analysis, peak data were examined by using GeneMapper v4.0 software (Applied Biosystems) to calculate the repeat number for each VNTR locus on the basis of fluorescent label and fragment length. The polymorphism index (discriminatory power) of individual or combined VNTR loci was calculated using GenAlEx v 6.5 (Peakall and Smouse, 2012). The diversity indices were classified into four groups based on geographical origin: North America, northwestern Europe, southeastern Europe and Oceania.

Performance criteria

The reproducibility was tested with a positive control in fragment analyses, performing repeated analyses of the same isolates and by comparing selected reference strains from different sources. Epidemiological concordance, the amount of different strains within an outbreak, was assessed on: (i) isolates from a single tree with several fire blight strikes and (ii) two subsets of isolates of an extensive outbreak in Slovenia in 2003 selected based on location and time of isolation. Stability of the selected markers was assessed by analysis of several reference *E. amylovora* strains stored in different laboratories and indirectly through repeated analysis of selected isolates.

Analysis of global E. amylovora *diversity*

The number of all alleles and the frequency distribution of repeats was determined using GenAlEx software (Peakall and Smouse, 2012). All MSTs were generated in Bionumerics 7.0 (Applied Maths, Sint-Martens-Latem, Belgium). The Bayesian clustering approach, implemented in the software STRUCTURE 2.4.3, was used to infer population structure and assign individuals to groups characterized by distinct allele frequencies (Pritchard *et al*., 2000). The method estimates a probability of ancestry for each individual from each of the groups. Individuals are assigned to one cluster or jointly to two or more clusters if their genotypes indicate that they were admixed. Twenty independent runs of STRUCTURE were performed by setting the number of subpopulations or groups (K) from 1 to 10, with 500 000 burn-in replicates and a run length of 750 000 replicates to decide which value of K best fits the data (Evanno *et al*., 2005). Clustering of isolates of *E. amylovora* was evaluated for the inferred number of groups. STRUCTURE was run using the admixture model without prior population information, which assumes correlated allele frequencies for our MLVA data. Nei's genetic identities (Nei, 1973) between groups identified in the clustering methods were computed in GenAlEx (Peakall and Smouse, 2012).

Analysis of regional E. amylovora *populations*

The datasets from Slovenia ($n = 461$) and Switzerland ($n = 108$) were analysed for molecular variance using AMOVA (Excoffier *et al*., 1992) in GenAlEx (Peakall and Smouse, 2012). Two distinct analyses were performed by grouping the isolates by host plant and by year of isolation. All data were handled as haploid-SSR and permutation test were performed with 999 random permutations of haplotypes. Correlation of geographic and genetic distance were assessed with Mantel test in GenAlEx (Peakall and Smouse, 2012). Geographic distance was handled untransformed, as well as log and ln transformed to cover all possibilities. Isolates from Slovenia include isolates from: (i) the first findings of fire blight in 2001 and 2002 in Gorenjska region (Knapic *et al*., 2004), (ii) isolates from two extensive outbreaks in 2003 and 2007 and (iii) isolates from intervening years up to 2009 when conditions were not very conducive for fire blight symptom development and disease occurred sporadically. Our collection of Swiss isolates includes isolates from the years 2002–2009.

Acknowledgements

The authors thank partners in the EU EUPHRESCO ERA-Net ERWINDECT and Jay Norelli, USDA-ARS-AFRS, Joseph Nemeth, Plant Protection and Soil Conservation Service of Baranya County, Hungary, Nataliya Drenova, Bacteriology laboratory, Federal State Enterprise 'All-Russian Plant Quarantine Centre' and Esther Moltman, Landesanstalt für Pflanzenschutz, Germany, for bacterial isolates of *Erwinia amylovora* or their DNA. We also thank Beatrice Frey and Markus Oggenfuss for technical assistance. Financial support was provided by the Swiss Federal Office of Agriculture (ACHILLES), the European Union FP7 KBBE (Q-Detect) and the EU EUPHRESCO ERA-Net (PhytFire) projects and the Biotechnology and Systems Biology of Plants programme (P4–0165). This work was conducted within the Swiss ProfiCrops and European COST Action 864 research networks.

Conflict of interest

The authors declare no conflict of interest.

References

Aandahl, R.Z., Reyes, J.F., Sisson, S.A., and Tanaka, M.M. (2012) A model-based Bayesian estimation of the rate of evolution of VNTR loci in *Mycobacterium tuberculosis*. *PLoS Comput Biol* **8:** e1002573.

Achtman, M. (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* **62:** 53–70.

Baltrus, D.A., Nishimura, M.T., Romanchuk, A., Chang, J.H., Mukhtar, M.S., Cherkis, K., *et al*. (2011) Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLoS Pathog* **7:** e1002132.

van Belkum, A. (1999) The role of short sequence repeats in epidemiologic typing. *Curr Opin Microbiol* **2:** 306–311.

van Belkum, A. (2007) Tracing isolates of bacterial species by multilocus variable number of tandem repeat analysis (MLVA). *FEMS Immunol Med Microbiol* **49:** 22–27.

van Belkum, A., Scherer, S., van Leeuwen, W., Willemse, D., van Alphen, L., and Verbrugh, H. (1997) Variable number of tandem repeats in clinical strains of *Haemophilus influenzae*. *Infect Immun* **65:** 5017–5027.

Bogdanove, A.J., Koebnik, R., Lu, H., Furutani, A., Angiuoli, S.V., Patil, P.B., *et al*. (2011) Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomona*s spp. *J Bacteriol* **193:** 5450–5464.

Bonn, W.G., and van der Zwet, T. (2000) Distribution and economic importance of fire blight. In *Fire Blight the Disease and Its Causative Agent, Erwinia amylovora*. Vanneste, J.L. (ed.). Wallingford, UK: CAB Intl, pp. 37–53.

Braun-Kiewnick, A., Altenbach, D., Oberhänsli, T., Bitterlin, W., and Duffy, B. (2011) A rapid lateral-flow immunoassay for phytosanitary detection of *Erwinia amylovora* and on-site fire blight diagnosis. *J Microbiol Methods* **87:** 1–9.

Bui Thi Ngoc, L., Vernière, C., Jarne, P., Brisse, S., Guérin, F., Boutry, S., *et al*. (2009) From local surveys to global surveillance: Three high-throughput genotyping methods for epidemiological monitoring of *Xanthomonas citri* pv. *citri* pathotypes. *Appl Environ Microbiol* **75:** 1173–1184.

Caporale, L.H. (2003) Natural selection and the emergence of a mutation phenotype: an update of the evolutionary synthesis considering mechanisms that affect genome variation. *Annu Rev Microbiol* **57:** 467–485.

Croucher, N.J., Harris, S.R., Fraser, C., Quail, M.A., Burton, J., van der Linden, M., *et al*. (2011) Rapid pneumococcal evolution in response to clinical interventions. *Science* **331:** 430–434.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* **12:** 499–510.

De Fonzo, V., Aluffi-Pentini, F., and Parisi, V. (2008) JSTRING: a novel Java tandem repeats searcher in genomic sequences with an interactive graphic output. *Open Appl Inform J* **2:** 14–17.

Denning, W. (1793) On the decay of apple trees. *NY Soc Prom Agric Arts Manuf Trans* **2:** 219–222. (Transactions of the Society for the Promotions of Useful Arts, in the State of New York).

Denoeud, F., and Vergnaud, G. (2004) Identification of polymorphic tandem repeats by direct comparison of genome sequence from different bacterial strains: a web-based resource. *BMC Bioinf* **5:** 4.

Donat, V., Biosca, E.G., Peñalver, J., and López, M.M. (2007) Exploring diversity among Spanish strains of *Erwinia amylovora* and possible infection sources. *J Appl Microbiol* **103:** 1639–1649.

Evanno, G., Regnaut, S., and Goudet, J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* **14:** 2611–2620.

Excoffier, L., Smouse, P.E., and Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131:** 479–491.

Feil, E.J., Holmes, E.C., Bessen, D.E., Chan, M.-S., Day, N.P.J., Enright, M.C., *et al*. (2001) Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci U S A* **98:** 182–187.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., *et al*. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:** 496–512.

Gehring, I., and Geider, K. (2012) Differentiation of *Erwinia amylovora* and *Erwinia pyrifoliae* strains with single nucleotide polymorphisms and by synthesis of dihydrophenylalanine. *Curr Microbiol* **65:** 73–84.

Geluk, F., Eijk, P.P., van Ham, S.M., Jansen, H.M., and van Alphen, L. (1998) The fimbria gene cluster of nonencapsulated *Haemophilus influenzae*. *Infect Immun* **66:** 406–417.

Girard, J.M., Wagner, D.M., Vogler, A.J., Keys, C., Allender, C.J., Drickamer, L.C., and Keim, P. (2004) Differential plague-transmission dynamics determine *Yersinia pestis* population genetic structure on local, regional, and global scales. *Proc Natl Acad Sci U S A* **101:** 8408–8413.

Gironde, S., and Manceau, C. (2012) Housekeeping gene sequencing and multilocus variable-number tandem-repeat analysis to identify subpopulations within *Pseudomonas syringae* pv. *maculicola* and *Pseudomonas syringae* pv. *tomato* that correlate with host specificity. *Appl Environ Microbiol* **78:** 3266–3279.

Gladieux, P., Zhang, X.-G., Ròldan-Ruiz, I., Caffier, V., Leroy, T., Devaux, M., *et al*. (2010) Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. *Mol Ecol* **19:** 658–674.

Harris, S.R., Feil, E.J., Holden, M.T.G., Quail, M.A., Nickerson, E.K., Chantratita, N., *et al*. (2010) Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327:** 469–474.

Hood, D.W., Deadman, M.E., Jennings, M.P., Bisercic, M., Fleischmann, R.D., Venter, J.C., and Moxon, E.R. (1996) DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *Proc Natl Acad Sci U S A* **93:** 11121–11125.

Jock, S., Donat, V., López, M.M., Bazzi, C., and Geider, K. (2002) Following spread of fire blight in Western, Central and Southern Europe by molecular differentiation of *Erwinia amylovora* strains with PFGE analysis. *Environ Microbiol* **4:** 106–114.

Jock, S., Wensing, A., Pulawska, J., Drenova, N., Dreo, T., and Geider, K. (2013) Molecular analyses of *Erwinia amylovora* strains isolated in Russia, Poland, Slovenia and Austria describing further spread of fire blight in Europe. *Microbiol Res* **168**. doi:10.1016/j.micres.2013.01.008.

Kim, W.S., and Geider, K. (1999) Analysis of variable short-sequence DNA repeats on the 29 kb plasmid of *Erwinia amylovora* strains. *Eur J Plant Pathol* **105:** 703–713.

Klevytska, A.M., Price, L.B., Schupp, J.M., Worsham, P.L., Wong, J., and Keim, P. (2001) Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J Clin Microbiol* **39:** 3179–3185.

Knapic, V., Potocnik, A., Skerlavaj, V., and Brecl, A. (2004) First outbreaks of fireblight in Slovenia. *EPPO Bull* **34:** 351–356.

Lindstedt, B.-A. (2005) Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis* **26:** 2567–2582.

Llop, P., Donat, V., Rodríguez, M., Cabrefiga, J., Ruz, L., Palomo, J.L., *et al*. (2006) An indigenous virulent strain of *Erwinia amylovora* lacking the ubiquitous plasmid pEA29. *Phytopathology* **96:** 900–907.

McGhee, G.C., and Sundin, G.W. (2012) *Erwinia amylovora* CRISPR elements provide new tools for evaluating strain diversity and for microbial source tracking. *PLoS ONE* **7:** e41706.

McManus, P.S., and Jones, A.L. (1995) Genetic fingerprinting of *Erwinia amylovora* strains isolated from tree-fruit crops and *Rubus* spp. *Phytopathology* **85:** 1547–1553.

Malachowa, N., Sabat, A., Gniadkowski, M., Krzyszton-Russjan, J., Empel, J., Miedzobrodzki, J., *et al*. (2005) Comparison of multiple-locus variable-number tandem-repeat analysis with pulsed-field gel electrophoresis, *spa* yyping, and multilocus sequence typing for clonal characterization of S*taphylococcus aureus* isolates. *J Clin Microbiol* **43:** 3095–3100.

Mann, R.A., Blom, J., Bühlmann, A., Plummer, K.M., Beer, S.V., Luck, J.E., *et al*. (2012) Comparative analysis of the Hrp pathogenicity island of *Rubus*- and Spiraeoideae-infecting *Erwinia amylovora* strains identifies the IT region as a remnant of an integrative conjugative element. *Gene* **504:** 6–12.

Mann, R.A., Smits, T.H.M., Bühlmann, A., Blom, J., Goesmann, A., Frey, J.E., *et al*. (2013) Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core. *PLoS ONE* **8:** e55644.

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., *et al*. (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol Plant Pathol* **13:** 614–629.

Mazars, E., Lesjean, S., Banuls, A.-L., Gilbert, M., Vincent, V., Gicquel, B., *et al*. (2001) High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc Natl Acad Sci U S A* **98:** 1901–1906.

Medini, D., Serruto, D., Parkhill, J., Relman, D.A., Donati, C., Moxon, R., *et al*. (2008) Microbiology in the post-genomic era. *Nat Rev Microbiol* **6:** 419–430.

Mutreja, A., Kim, D.W., Thomson, N.R., Connor, T.R., Lee, J.H., Kariuki, S., *et al*. (2011) Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* **477:** 462–465.

Nei, M. (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci U S A* **70:** 3321–3323.

Parkhill, J., and Wren, B. (2011) Bacterial epidemiology and biology – lessons from genome sequencing. *Genome Biol* **12:** 230.

Peakall, R., and Smouse, P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* **28:** 2537–2539.

Pirc, M., Ravnikar, M., Tomlinson, J., and Dreo, T. (2009) Improved fireblight diagnostics using quantitative real-time PCR detection of *Erwinia amylovora* chromosomal DNA. *Plant Pathol* **58:** 872–881.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Rezzonico, F., Smits, T.H.M., and Duffy, B. (2011) Diversity, evolution and functionality of CRISPR regions in the fire blight pathogen *Erwinia amylovora*. *Appl Environ Microbiol* **77:** 3819–3829.

Rezzonico, F., Braun-Kiewnick, A., Mann, R.A., Rodoni, B., Goesmann, A., Duffy, B., and Smits, T.H.M. (2012) Lipopolysaccharide biosynthesis genes discriminate between *Rubus*- and Spiraeoideae-infective genotypes of *Erwinia amylovora*. *Mol Plant Pathol* **13:** 975–984.

Rico, A., Ortiz-Barredo, A., Ritter, E., and Murillo, J. (2004) Genetic characterization of *Erwinia amylovora* strains by amplified fragment length polymorphism. *J Appl Microbiol* **96:** 302–310.

Schnabel, E.L., and Jones, A.L. (1998) Instability of a pEA29 marker in *Erwinia amylovora* previously used for strain classification. *Plant Dis* **82:** 1334–1336.

Smith, J.M., Smith, N.H., O'Rourke, M., and Spratt, B.G. (1993) How clonal are bacteria? *Proc Natl Acad Sci U S A* **90:** 4384–4388.

Smits, T.H.M., Rezzonico, F., Kamber, T., Blom, J., Goesmann, A., Frey, J.E., and Duffy, B. (2010) Complete genome sequence of the fire blight pathogen *Erwinia amylovora* CFBP 1430 and comparison to other *Erwinia* spp. *Mol Plant Microbe Interact* **23:** 384–393.

Spratt, B. (2004) Exploring the concept of clonality in bacteria. In *Genomics, Proteomics, and Clinical Bacteriology*. Woodford, N., and Johnson, A. (eds). Clifton, NJ, USA: Humana Press, pp. 323–352.

Spratt, B.G., Hanage, W.P., and Feil, E.J. (2001) The relative contributions of recombination and point mutation to the diversification of bacterial clones. *Curr Opin Microbiol* **4:** 602–606.

Stockwell, V.O., and Duffy, B. (2012) Use of antibiotics in plant agriculture. *Rev Sci Tech* **31:** 199–210.

Stratilo, C.W., and Bader, D.E. (2012) Genetic diversity among *Bacillus anthracis* soil isolates at fine geographic scales. *Appl Environ Microbiol* **78:** 6433–6437.

Supply, P., Warren, R.M., Bañuls, A.-L., Lesjean, S., Van Der Spuy, G.D., Lewis, L.-A., *et al*. (2003) Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. *Mol Microbiol* **47:** 529–538.

Thomson, S.V. (2000) Epidemiology of fire blight. In *Fire Blight the Disease and Its Causative Agent, Erwinia amylovora*, Vanneste, J.L. (ed.). Wallingford, UK: CAB Intl. pp. 9–35.

Torres-Cruz, J., and van der Woude, M.W. (2003) Slipped-strand mispairing can function as a phase variation mechanism in *Escherichia coli*. *J Bacteriol* **185:** 6990–6994.

U'Ren, J.M., Hornstra, H., Pearson, T., Schupp, J.M., Leadem, B., Georgia, S., *et al*. (2007) Fine-scale genetic diversity among *Burkholderia pseudomallei* soil isolates in Northeast Thailand. *Appl Environ Microbiol* **73:** 6678–6681.

Vogler, A.J., Keys, C., Nemoto, Y., Colman, R.E., Jay, Z., and Keim, P. (2006) Effect of repeat copy number on variable-number tandem repeat mutations in *Escherichia coli* O157:H7. *J Bacteriol* **188:** 4253–4263.

Vogler, A.J., Keys, C.E., Allender, C., Bailey, I., Girard, J., Pearson, T., *et al*. (2007) Mutations, mutation rates, and evolution at the hypervariable VNTR loci of *Yersinia pestis*. *Mutat Res* **616:** 145–158.

Walsh, P.S., Metzger, D.A., and Higuchi, R. (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10:** 506–513.

Wang, D., Korban, S.S., and Zhao, Y. (2010) Molecular signature of differential virulence in natural isolates of *Erwinia amylovora*. *Phytopathology* **100:** 192–198.

Zaleski, P., Wojciechowski, M., and Piekarowicz, A. (2005) The role of Dam methylation in phase variation of *Haemophilus influenzae* genes involved in defence against phage infection. *Microbiology* **151:** 3361–3369.

Zhang, Y., and Geider, K. (1997) Differentiation of *Erwinia amylovora* strains by pulsed-field gel electrophoresis. *Appl Environ Microbiol* **63:** 4421–4426.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Haploid allele frequencies of strains sampled in Slovenia for different outbreak years (A) and for different host plants (B) Bars show frequencies of corresponding VNTR alleles, subdivided into different VNTR loci.

Fig. S2. Geographical distribution of *Erwinia amylovora* during epidemic outbreaks of 2003 (A) and 2007 (B) in Slovenia, colored by VNTR type. The size of pins correlates to the genetic distance from the first haplotype detected in 2002. Grey pins show outlier haplotypes only detected once.

Table S1. Test panel of *Erwinia amylovora* strains used in development of MLVA, their MLVA profiles, CRISPR types and STRUCTURE groups. For references to the individual strains: see reference Rezzonico *et al.*, 2011.

Table S2. Different species used for specificity testing of *Erwinia amylovora* VNTRs. Only Erwinia sp. 223B was scored on all six loci displaying a unique haplotype and thus proving specificity of the developed MLVA typing scheme.

Table S3. Panel of single tree isolates from Switzerland and their MLVA profile showing multiple VNTR types on every tree assayed except tree number two.

Table S4. List of VNTRs, primers used for amplification of repeat regions and their basic characteristics.

Table S5. List of all *Erwinia amylovora* isolates included in the analysis and their allele profiles. This list includes, where available, the host plant it is isolated from and the year of isolation.