MR. DAMIEN RICHARD (Orcid ID : 0000-0001-9179-7315)

DR. OLIVIER PRUVOST (Orcid ID : 0000-0002-3175-9795)

DR. FRANÇOIS BALLOUX (Orcid ID : 0000-0003-1978-7715)

DR. ADRIEN RIEUX (Orcid ID : 0000-0002-7221-0010)

Article type : Original Article

Time-calibrated genomic evolution of a monomorphic bacterium during its establishment as an endemic crop pathogen

Running title: Genomic evolution of a pathogenic bacterium

Authors

Damien Richard[a,b,c,#]; Olivier Pruvost[a]; François Balloux[d]; Claudine Boyer[a]; Adrien Rieux[a]; Pierre Lefeuvre[a]

Affiliation

[a] Cirad, UMR PVBMT, F-97410 St Pierre, Réunion, France.

[b] ANSES, Plant Health Laboratory, F-97410 St Pierre, Réunion, France.

[c] Université de la Réunion, UMR PVBMT, F-97490 St Denis, Réunion, France.

[d] UCL Genetics Institute, University College London, London WC1E 6BT, UK.

[#] Corresponding author: Damien Richard, richarddamienfr@gmail.com

Competing interests: The authors declare that they have no competing interests.

Keywords

Genomic evolution; Dated phylogeny; Citrus canker; Gene turnover rate; nucleotide substitution rate

Abstract

Horizontal gene transfer is of major evolutionary importance as it allows for the redistribution of phenotypically important genes among lineages. Such genes with essential functions include those involved in resistance to antimicrobial compounds and virulence factors in pathogenic bacteria. Understanding gene turnover at microevolutionary scales is critical to assess the pace of this evolutionary process. Here we characterized and quantified gene turnover for the epidemic lineage of a bacterial plant pathogen of major agricultural importance worldwide. Relying on a dense geographic sampling spanning 39 years of evolution, we estimated both the dynamics of single nucleotide polymorphism accumulation and gene content turnover. We identified extensive gene content variation among lineages even at the smallest phylogenetic and geographic scales. Gene turnover rate exceeded nucleotide substitution rate by three orders of magnitude. Accessory genes were found preferentially located on plasmids, but we identified a highly plastic chromosomal region hosting ecologically important genes such as transcription activator-like effectors. Whereas most changes in the gene content are probably transient, the rapid spread of a mobile element conferring resistance to copper compounds widely used for the management of plant bacterial pathogens illustrates how some accessory genes can become ubiquitous within a population over short time-frames.

Introduction

Biological invasions represent a major threat to human, animal and plant health as well as to the preservation of global biodiversity (Guillemaud, Ciosi, Lombaert, & Estoup, 2011). Geographical expansion of pre-adapted pests and pathogens together with climate change are major drivers of the emergence of plant diseases (Anderson et al., 2004). Such expansions increased in frequency over the 20th century due to the marked intensification of international trade and human transportation through globalization (Hulme, 2009). Emergence can also be the consequence of bacterial adaptation, allowing for example host jump or resistance to antimicrobial compounds used for disease management (Engering, Hogerwerf, & Slingenbergh, 2013). Such new traits can be the consequence of point mutations (e.g., streptomycin resistance in *Erwinia amylovora*, the bacterium responsible for fireblight of pear and apple trees) or, more often, of gene content variations (Jacques et al., 2016; McManus, Stockwell, Sundin, & Jones, 2002).

Diverse mechanisms are associated with variation in gene content: whereas gene loss can be caused by mutational deletion, phage-mediated excision or homologous recombination, gene gain can occur through gene duplication or horizontal gene transfer (Lerat, Daubin, Ochman, & Moran, 2005; Rocha, 2008). Whereas genes seem to be readily sampled from the accessory gene pool and transferred between lineages, it appears that most gene content changes are only transient (Croucher et al., 2014; Hall, Brockhurst, & Harrison, 2017). Theoretical (Haegeman & Weitz, 2012; Vos, Hesselman, Te Beek, van Passel, & Eyre-Walker, 2015) and empirical (Andreani, Hesse, & Vos, 2017; Hao & Golding, 2006; Zhou et al., 2018) studies suggest a neutral or deleterious effect for most gene content changes. Nonetheless, gain and loss of accessory genes is known to play an important role in bacterial adaptation (Hall et al., 2017). Indeed, accessory genes can provide access to novel functions, such as enhanced in-host multiplication, the ability to colonize a new ecological niche, interspecies bacterial competition or antimicrobial resistance (Barash & Manulis-Sasson, 2009; Kado, 2015; McManus et al., 2002; Pilla & Tang, 2018; Swarup, De Feyter, Brlansky, & Gabriel, 1991; Thynne, McDonald, & Solomon, 2015; Weber, Ly, Irwin, Pukatzki, & Feldman, 2015).

With the increasing easiness of complete or nearly complete genomic sequence production, analysis of large genomic datasets has allowed the reconstruction of the evolutionary history of bacterial pathogens (Estoup & Guillemaud, 2010; Ruh, Briand, Bonneau, Jacques, & Chen, 2017). One step further, the reconstruction of ancestral gene content in phylogenetic studies has shown promise for understanding the emergence, adaptation, geographical expansion and host specificity of plant pathogenic bacteria (Chen et al., 2018; Monteil et al., 2016; Vinatzer, Monteil, & Clarke, 2014).

Among plant pathogenic bacteria, the genus *Xanthomonas* groups pathologically highly specialized infraspecific lineages causing some internationally major plant diseases and collectively infecting a broad range of plants that are pivotal to food security, cash crops and wild plant species (Jacques et al., 2016). Many of these lineages are regarded as 'monomorphic', i.e., bacteria that display low genetic diversity and nucleotide substitution rates (Achtman, 2012). Adaptation through acquisition of genes by horizontal transfer has been reported for several xanthomonads (Behlau, Hong, Jones, & Graham, 2013; Chen et al., 2018; El Yacoubi, Brunings, Yuan, Shankar, & Gabriel, 2007; Lima, Paquola, Varani, Van Sluys, & Menck, 2008). The monomorphic bacterium *Xanthomonas citri* pv. *citri* causes Asiatic citrus canker, an economically important disease that threatens citrus industries in most areas of production. Despite restricted levels of sequence divergence between strains, *Xanthomonas citri* pv. *citri* genome was reported to display variations in some horizontally acquired regions integrated into the chromosome (i.e., genomic islands) (Gordon et al., 2015) or plasmid content (Gochez et al., 2018; Richard, Ravigne, et al., 2017). As such, plasmids were suggested to be mobile genetic elements of prime importance in the evolution of *Xanthomonas* species (Halary, Leigh, Cheaib, Lopez, & Bapteste, 2010). Indeed, within this genus some key phenotypes are encoded by plasmid genes, of which some type III transcription-activator like effector (TALE) genes (da Silva et al., 2002; Ruh et al., 2017). Essential for pathogenicity, TALEs are injected in plant host cells in which they activate the expression of eukaryotic genes that promote pathogenicity or induce defense (Ryan et al., 2011). Similarly, plasmid-borne resistance to antimicrobial compounds (antibiotics and heavy metals) were reported in the genus *Xanthomonas* (Gardan, Brault, & Germain, 1993; Hyun, Kim, Yi, Hwang, & Park, 2012; Minsavage, Canteros, & Stall, 1990; Niu et al., 2015; Richard, Ravigne, et al., 2017).

In the present study, we undertook a population genomics analysis of strains sampled over 39 years with the aim of understanding the phylogeography of the pathogen at a regional scale and the respective effect of gene content variation and nucleotide substitutions on genome evolution. We further investigated how the pathogen adapted to copper, a novel and recurrent selection pressure brought up through disease management. More specifically, we addressed the following questions: What was the timeframe of *Xanthomonas citri* pv. *citri* establishment in the South West Indian Ocean islands? What was the rate of transfer of the copper resistant plasmid within populations in Réunion (i.e., the island where this plasmid was detected)? What is the importance of genes shuttled by plasmids in the composition of the accessory genome of an outbreak *Xanthomonas* population? Besides *Xanthomonas citri* pv. *citri* evolution, we propose hypotheses on the introduction history of *Xanthomonas citri* pv. *citri* in the South West Indian Ocean region, going back to its first probable introduction ca. two centuries ago. We highlighted a very high ratio of gene turnover (i.e. the combination of gene gain and gene loss (Graña-Miraglia et al., 2017; Nowell, Green, Laue, & Sharp, 2014)) to

nucleotide substitution rate and emphasized that most gene content changes may be short-lived. We also emphasized the significance of plasmids and a newly described chromosomal genomic island as reservoirs of accessory genes in the analyzed *Xanthomonas citri* pv. *citri* population.

Results

Global phylogeny

The genomic analysis of 284 strains representing worldwide *Xanthomonas citri* pv. *citri* diversity (Figure 1, Table S1) revealed 7,005 high confidence SNPs from which a phylogeny was reconstructed (Supplementary Figure S1A). Within the phylogeny, all the 20 strains from the Northern part of the Indian Ocean (Maldives and Seychelles) fell into four of the numerous unresolved clades. Conversely, a well-resolved monophyletic clade comprising 221 strains included all the 210 South West Indian Ocean strains along with eleven strains from Martinique (French West Indies). In this clade, referred to hereafter as the South West Indian Ocean clade, subclades tended to group strains sampled from a single island (Figure 2). For Réunion, the sole island where a dense sampling of geo-located groves was performed, no obvious spatial structure was apparent and all but one of the citrus groves hosted strains from multiple Réunion subclades (Supplementary Figure S1B).

Temporal signal and time calibration

Although there was no significant temporal signal (i.e., relationship between time and genetic distance) for the whole worldwide phylogeny, both the root-to-tip regression test (slope: 3.2×10^{-5}, p-value: 2.3×10^{-4} and R^2: 0.06) and the date-randomization test (no overlap between inferences performed on real vs date-randomized datasets) revealed the presence of temporal signal within the South West Indian Ocean clade (Supplementary Figure S2). Importantly, our results showed no evidence of confounding between temporal and genetic structures (Mantel test: $r = 0.028$, p-value = 0.188), which suggested that the temporal signal detected is reliable and robust (Murray et al., 2016).

In order to reconstruct and time scale the evolutionary history of *Xanthomonas citri* pv. *citri*, a phylodynamic analysis was conducted using the BEAST framework (Drummond & Rambaut, 2007) on the 221 strains of the South West Indian Ocean clade (3,403 SNPs over a 5,175,554 bp alignment), using an exponential growth tree model and an uncorrelated lognormal relaxed clock (estimated standard deviation: 0.26 (95% HPD 0.18-0.34)). Overall, the Maximum Likelihood and the Bayesian trees had a similar topology (Supplementary Figure S3). The South West Indian Ocean tree root age was established at around 1818 (95% HPD: 1762-1868, Table 1), with an estimated substitution rate of 8.4×10^{-8} substitutions per site per year (95% HPD: 6.9×10^{-8}-1.0×10^{-7}). This corresponds to 0.43 SNP per genome per year (95% HPD: 0.35-0.51).

Accessory genome content

In order to test our ability to reconstruct the gene content of *Xanthomonas citri* pv. *citri* strains, two strains were sequenced in triplicate and used as input of the in-house gene content inference pipeline. The results we obtained were compared to those of obtained from a pan-genome estimation pipeline based on *de novo* assemblies (Supplementary Figure S4). From the in-house pipeline, the gene content (4,903 genes) was identical between all replicates for one of the strains, while five genes out of 4,852 were not detected in at least one replicate of the other strain, with a pairwise maximum difference between replicates of two genes. Importantly the five genes were not clustered within the genomes. It is therefore likely that events (i) involving the gain or loss of numerous genes or (ii) displaying a phylogenetic signal (i.e. when the same event is revealed in closely related individuals) are genuine. Conversely, from 30 to 134 genes were differentially detected depending on the conditions when using the pan-genome reconstruction pipeline based on *de novo* assemblies only. Importantly, the gene detection pipeline we employed mostly relies on the mapping of short reads on a reference rather than solely on short read assembly. This procedure was employed for gene detection because short read assemblies frequently result in fragmented genomes and truncated genes that impairs pan-genome reconstruction of some of the most frequently used pan-genome estimation programs (Fouts, Brinkac, Beck, Inman, & Sutton, 2012; Page et al., 2015). Conversely, the mapping procedure implemented here suffers of a lack of positional information, the inability to infer pseudogenization, paralogy or copy number variation. Our gene detection protocol will therefore be more exhaustive in terms of core-genome completion but overly conservative in terms of the number of genes per strain.

After *de novo* genome assembly, gene detection and gene clustering of the 221 South West Indian Ocean clade strains, a pan-genome size of 5,046 genes was identified. The core genome comprised 4,347 genes (i.e., detected in all strains), while the accessory genome comprised 699 genes (i.e., absent from at least one strain). Of these accessory genes, 336 were assigned to the chromosome, 339 to plasmids and 24 remained unassigned (Supplementary Figure S5). Most of the accessory genes (70%) did not match any known Cluster of Orthologous Group (COG); 4% matched replication, recombination and repair functions; 3% inorganic ion transport and other COG categories each comprised fewer than 3% of the gene clusters. As shown after the gene content analysis of sequencing replicates, the gene detection pipeline may have limitations associated with the detection of unique genes that have been gained or lost in single strains. Excluding the fourteen out of 699 genes that displayed such characteristics, we then used a Bayesian reconstruction of the ancestral states in terms of gene presence/absence to estimate gene gain, loss and total substitution rates for both the plasmid-borne and the chromosome-borne genes. Importantly, while the rates of turnover are reported in number of

gene changes per site per year, these represent large over-estimates of the genuine rates of events involved in gene turnover: multiple genes are frequently gained or lost in single recombination events (Bennett, 2004; Norman, Hansen, & Sorensen, 2009; Osborn & Böltner, 2002) and rates of turnover of neighbouring genes are thus most probably not independent.

The plasmid gene turnover rate ($2.24x10^{-3}$ gene gains or losses per gene and per year 95% HPD: $1.73x10^{-3}$-$3.42x10^{-3}$) appeared to be significantly higher than that of the chromosome ($8.83x10^{-5}$, 95% HPD: $8.14x10^{-5}$–$9.67x10^{-5}$). Nevertheless, at the replicon scale, these rates largely overlapped with 0.35 gene gains or losses per genome and per year for plasmid-borne genes (95% HPD 0.27–0.53) and 0.40 for chromosome-borne genes (95% HPD 0.37–0.44). Chromosome gene gain and loss rates were in the same order of magnitude (gene gain rate $1.37x10^{-5}$, 95% HPD $9.83x10^{-6}$-$1.80x10^{-5}$ and gene loss rate $7.45x10^{-5}$, 95% HPD $7.05x10^{-5}$-$7.89x10^{-5}$, respectively). Similar results were observed for the plasmid-borne genes (gene gain rate $1.42x10^{-3}$, 95% HPD $1.15x10^{-3}$-$2.04x10^{-3}$ and gene loss rate $8.15x10^{-4}$, 95% HPD $5.16x10^{-4}$-$1.41x10^{-3}$). For both plasmid and chromosome, the highest gene turnover rates were associated with external branches, but no significant difference of rates were detected depending on the branch ages (Supplementary Figure S6, Table 1).

Gain, loss and mosaic structure of plasmids

Whereas it was not possible to achieve full plasmid reconstruction using our short-read and gene-based approach, we were able to assess the presence of the gene content associated to the known set of *Xanthomonas citri* pv. *citri* plasmids (Richard, Ravigne, et al., 2017) in each strain. We were able to define the presence of four distinct plasmid groups with up to five distinct alleles associated with the South West Indian Ocean *Xanthomonas citri* pv. *citri* strains (see the right side of Figure 2). Importantly, these alleles were confirmed with long-read sequencing and assembly on a subset of strains (Supplementary Table S2).

Asiatic citrus canker control largely relies on the repeated applications of copper-based pesticides (Graham, Gottwald, Cubero, & Achor, 2004). In Réunion over the last decade, copper resistance has emerged with the integration from an unknown source of a mobile heavy metal resistance plasmid (pCu^R) in previously established *Xanthomonas citri* pv. *citri* strains (Richard, Ravigne, et al., 2017). Interestingly, the pCu^R gene content was identical for its 42 occurrences (31 in Réunion and 11 in Martinique). From the phylogenetic inference, the mean number of full pCu^R plasmid gain and loss were 7 (95%HPD: 6 - 9) and 1 (95%HPD: 0 - 3), respectively. Most of the pCu^R-bearing strains (out of a total of 37) formed a monophyletic group within subclade 6, the remaining five strains, originating from Réunion, were distributed in three distinct subclades (Figure 2). Fully identical pCu^R plasmids (without any SNP) were carried by distantly related strains isolated both in citrus groves and a commercial nursery (Supplementary Figure S7). In contrast with the dispensable

character of pCu^R, all *Xanthomonas citri* pv. *citri* strains known to date carry a pathogenicity-related plasmid gene set, organized as one or multiple plasmids (a few of which have been described so far (da Silva et al., 2002; Gochez et al., 2018).

Xanthomonas citri pv. *citri* pXac64-like plasmids (da Silva et al., 2002) typically carry *pthA4*, a transcription activator-like effector (TALE) gene required to produce citrus canker symptoms (Swarup et al., 1991). As expected, a pXac64-like plasmid was found in every South West Indian Ocean clade strain. Five distinct plasmid variants were detected. In particular, 49 strains from Mauritius, Réunion and Rodrigues displayed a deletion of 26 genes mostly coding for plasmid transfer and maintenance functions (Figure 2). The plasmid pLJ207-7.3 is closely related to pXac47 (Martins, Machado, Silva, Takita, & de Souza, 2016) but comprises a TALE gene whereas pXac47 does not. The functions of the proteins of these plasmids were mostly unknown (24 proteins), but included plasmid maintenance and transfer (17 proteins), as well as other functions (15 proteins). While most strains carried a pLJ207-7.3-like plasmid, it was absent in 77 South West Indian Ocean clade strains (Figure 2). Of these, 61 were phylogenetically clustered and corresponded to all but one individual of clade 6, a clade that included most of the Cu^R strains. Sixteen other strains lacking the plasmid were scattered throughout the phylogeny and had been sampled in Réunion, Mauritius, Mohéli, Mayotte and Rodrigues. As for pCu^R, both pXac64 and pXac47 variants were dispersed in the phylogeny. The geographic distribution of the plasmid alleles did not display any clear structure. Indeed, we observed up to six distinct plasmid profiles in a single grove in Réunion.

A 39.8kb plasmid including 40 genes was found in a single strain from Mauritius. The existence of the plasmid was confirmed through long-read sequencing. No other strain presented these genes. Whereas this plasmid has never been reported in *Xanthomonas citri* pv. c*itri*, or in any South West Indian Ocean bacterial strain, similar plasmids (nucleotide identity >75%) were previously described in *Xylella fastidiosa* (GenBank accession CP014330.2) and *Xanthomonas oryzae* (GenBank accession CP007810.1) isolated in Argentina and China, respectively. Besides conjugation and plasmid maintenance, no specific function could be associated to the proteins encoded on the plasmid.

A highly polymorphic chromosomal region

Compared to plasmids, *Xanthomonas citri* pv. *citri* chromosomal gene content was rather homogeneous with only 336 accessory genes. Using gene positions of a high-quality circular chromosome sequence (*Xanthomonas citri* pv. *citri* LH276 isolated from Réunion, GenBank accession CP018854.1) as a reference, we located 118 South West Indian Ocean chromosomal accessory genes in one genomic region, constituting a genomic island. The remaining 218 genes were dispersed along the chromosome in groups of up to seven genes or had no

homologue in the chosen reference. Long-read based *de novo* assemblies of 13 strains confirmed extensive gene variations of the genomic island, but also demonstrated that the chromosome was otherwise not rearrangement-prone (only one strain displayed a 1.2Mb inversion when compared to the LH276 reference). The genomic island carried between 32 to 92 genes, bordered by a phage integrase and a transposase on the 5' side and by one or multiple incomplete *AttR* sites on the 3' side (Figure 3). No deviation in GC content was detected for this region (data not shown). In most strains, the genomic island carried putative multidrug efflux pumps and a class C beta-lactamase was detected in five out of 13 strains. Most notably, the genomic island comprised one or two usually plasmid-borne TALE genes in seven strains (Figure 3). All chromosomal TALE genes were surrounded on either one or both sides by inverted repeats nucleotide motifs highly similar (>80% nucleotide identity) to that of the known Tn*3*-like *Xanthomonas* transposon Tn*Xax*1 (Ferreira et al., 2015). Additionally, a Tn*3* transposase was bordered by two TALE genes in the chromosomal genomic island of two strains. Chromosomal TALEs harboured between 8.5 and 28.5 variable di-residue motifs. Due to the repetitive nature of TALE genes, only long reads that covered the entire gene were used to generate the TALE sequence assemblies, therefore reducing the coverage of this region and impairing precise repeat variable di-residue motif determination (Streubel, Blucher, Landgraf, & Boch, 2012).

Discussion

In this study, we applied a population genomic approach on a set of bacterial strains obtained from an epidemic clade of *Xanthomonas citri* pv. *citri*, the causative agent of Asiatic citrus canker. We reconstructed its evolutionary history from its introduction in the South West Indian Ocean region to the present. We conjointly inferred the nucleotide substitution and gene turnover rates and found that gain and loss of accessory genes in *Xanthomonas citri* pv. *citri* is pervasive, even at the narrowest spatiotemporal scale.

Genetic structure within the South West Indian Ocean

Contrasting with the genetic diversity found in the northern Indian Ocean and India, South West Indian Ocean strains grouped in a single clade in the worldwide phylogeny. This monophyly indicates that the strains first introduced in the area were closely related, genetically and thus most probably geographically. It also suggests the absence of recent introductions from remote countries. We observed a strong geographic structure within the South West Indian Ocean. However, Réunion and Mauritius *Xanthomonas citri* pv. *citri* populations were both polyphyletic, impairing the identification of the location where the pathogen first established. This polyphyly could originate from a polyclonal primary inoculum, a frequent case for *Xanthomonas citri* pv. *citri* (Leduc et al., 2015). It could also indicate ancient exchanges of strains among islands. Indeed, the strain phylogeny suggests recent inter-island migration events among (i) Mauritius and Rodrigues (a remote Mauritian territory), (ii) the four islands in the Comoros Archipelago and (iii) Réunion and Martinique. In all three cases, political and/or economic links between islands likely represented conditions favourable to the long-range dispersal of *Xanthomonas citri* pv. *citri* through human-associated dispersal of contaminated citrus material (Leduc et al., 2015).

In contrast with the strong geographic structure of *Xanthomonas citri* pv. *citri* among islands, the analysis of strain diversity obtained after dense sampling in Réunion revealed little concordance between phylogeny and grove location, possibly reflecting the multiplicity of inoculum inputs during the lifespan of citrus groves. While grove contamination can occur via infected nursery plants when the grove is first established, it also occurs during periods of wind and rain (i.e., storms or hurricanes regularly hit islands in the region) and grove maintenance operations (Graham et al., 2004).

Emergence of *Xanthomonas citri* pv. *citri* in the South West Indian Ocean area

As we used the age of the ancestor of all the strains as a proxy to determine the age of the first introduction of *Xanthomonas citri* pv. *citri* in the South West Indian Ocean area, our estimate likely represents an upper bound: in the likely case where the present population results from the introduction of a population rather than

of a single ancestral strain, the age of the most recent common ancestor of the introduced population is expected to predate the introduction itself. This introduction was dated to 1818 (95% HPD: 1762-1868) and predates the earliest report of the disease in the area (1917 in Mauritius, (Aubert, 2014)). *Xanthomonas citri* pv. *citri* and its main host genus, *Citrus*, originated in Asia (Pruvost et al., 2014; Wu et al., 2018) and were probably spread beyond their area of origin by human-mediated movements of plants or plant propagative material. The time frame suggested here may be in line with two main hypotheses regarding the origin of the pathogen in the South West Indian Ocean.

The first hypothesis implies the introduction of contaminated plants by the French botanist and colonial administrator Pierre Poivre (1719-1786) thought to be at the root of the first introduction of citrus in the Mascarene Archipelago from several Asian countries in the mid-18th century (Du Pont de Nemours, 1797). The second hypothesis is that after the abolition of slavery in Mauritius and Réunion in 1835 and 1848, respectively, hundreds of thousands of indentured labourers from several Asian countries (the majority from India) were brought in to boost the agricultural workforce (Campbell, 1923; Carter, 2002; Govindin, 1994). This active flow of goods and people from the Asian continent may have led to the concomitant introduction of *Xanthomonas citri* pv. *citri* in the South West Indian Ocean area. Indeed, the emergence of pre-adapted plant pathogens due to migratory events is a well-recognized phenomenon (Anderson et al., 2004; McDonald & Stukenbrock, 2016; Yoshida et al., 2013). We were unable to accurately identify the geographic origin of the strains that first migrated to the South West Indian Ocean region. More strains from the hypothetical Asian cradle of *Xanthomonas citri* pv. *citri* will have to be analysed to provide a definite answer regarding the geographic origin of the pathogen and its migratory history.

Rates of evolution: nucleotide substitution and gene turnover rates

Because most horizontal gene transfer events are usually deleterious and short-lived (Gogarten & Townsend, 2005; Nowell et al., 2014), Vos et al. (2015) emphasized the need for studies explicitly quantifying the rate at which accessory genes are integrated and lost from datasets composed of closely-related bacterial strains. Consistent with theoretical expectations by Vos et al. (2015) and Iranzo, Wolf, Koonin, and Sela (2019), our results clearly suggest that the ratio of gene content changes to nucleotide substitution was indeed very high at the narrow evolutionary scale analysed herein. We estimated that *Xanthomonas citri* pv. *citri* nucleotide substitution rate comes within the lower range of known bacterial rates, consistent with the low levels of diversity in its core genome (Duchene et al., 2016). Conversely, chromosome and plasmid gene turnover rate estimates (expressed in number of gain or lost genes per gene and per year) were substantial with values of 8.83×10^{-5} and 2.24×10^{-3}, respectively.

We also report a high heterogeneity in gene turnover rates among lineages at a narrow phylogenetic scale, that seem to stabilize in deeper phylogenetic branches. It is important to note that the turnover rates we computed are overestimates of the rates at which genomic tracts are inserted or deleted. For example, the gain of a single large plasmid would represent the addition in a single step of a large number of genes. Also, the actual plasmid gene turnover rate on plasmids may be imperfectly captured by our analyses since we constrained the rates of plasmid evolution, a horizontally transferable genomic element, with that of the bacterial host. When expressed on a per-site basis the gene turnover rate was three orders of magnitude higher than nucleotide substitution rates. The high gene turnover rate estimates inferred herein, as compared to previous studies that reported nucleotide substitution and gene turnover rates in the same order of magnitude (Hao & Golding, 2006; Pradeep Reddy Marri, Hao, & Golding, 2006; P. R. Marri, Hao, & Golding, 2007; Touchon et al., 2009) likely reflects the high genetic relatedness among strains in our dataset and thus consequent ease of horizontal transfer. Whether such a global high gene turnover is neutral or driven by positive selection of transiently beneficial accessory genes among numerous quickly purged neutral or sightly deleterious genes, remains a debate (Andreani et al., 2017; Bolotin & Hershberg, 2017; Lobkovsky, Wolf, & Koonin, 2013; Vos et al., 2015; Zhou et al., 2013).

Transfer and maintenance of adaptive genes

In the South West Indian Ocean lineage, pCu^R is a good example of a recently acquired and successfully maintained mobile genetic element. Whereas these resistance genes were not detected in an extensive strain collection from Réunion before 2010, now copper-resistant strains are circulating widely. The presence of fully identical pCu^R plasmids in phylogenetically distant strains further confirms previous data and *in vitro* tests that demonstrated the mobility of pCu^R. Additionally, the loss rate of pCu^R was found to be seven times lower than its gain rate, suggesting that pCu^R is crucial to *Xanthomonas citri* pv. *citri* in the context of the repeated copper applications employed in Réunion citriculture. pCu^R stability can also reflect the functioning of plasmid maintenance systems such as toxin-antitoxin or partitioning.

The uniformity of pCu^R in terms of Single Nucleotide Polymorphism and gene content contrasts with its mosaic nature on a broader geographic and phylogenetic scale (Richard, Ravigne, et al., 2017). This suggests that pCu^R plasmids from Réunion were recently gained in a unique event (Richard, Tribot, et al., 2017) and have spread rapidly through the trade of contaminated citrus nursery plants. In this regard, tracking the future trajectory of pXac39, a previously uncharacterized plasmid observed in a single strain, would be particularly interesting. Overall, plasmids appeared much more prone to gene gain or loss than chromosomes, with a gene turnover rate two orders of magnitude higher. Consistent with earlier network analyses (Halary, Leigh, Cheaib,

Lopez, & Bapteste, 2010), plasmids represent a privileged type of vehicle for accessory genes in *Xanthomonas citri* pv. *citri.*

In opposition to plasmids, chromosomes of *Xanthomonas citri* pv. *citri* were found to be relatively homogeneous in terms of their gene content and synteny. Nevertheless, a genomic island with a highly variable gene content and organization was present in all strains (from distinct subclades) that were submitted to long-read sequencing. The genomic island was located downstream to a phage integrase, carried possible traces of phage integration (the *AttR* sites in its 3' end (Fogg, Colloms, Rosser, Stark, & Smith, 2014)) and signatures of Tn*3*-like transposons. Interestingly, some allelic versions of the genomic island comprised putatively adaptive genes, including a heavy metal efflux pump, an antibiotic resistance gene and up to two type III effectors of the TALE family.

Bacterial TALE genes are crucial for pathogenicity. This family of genes codes for proteins that trigger the activation of specific plant genes and modulate infection. While one of these genes (*pthA4*) is recognized as the major pathogenicity determinant of *Xanthomonas citri* pv. *citri* pathotype A, their genomes always contain several alternative TALE genes, some of which were found to modify its virulence (Roeschlin et al., 2019; Shiotani, Fujikawa, Ishihara, Tsuyumu, & Ozaki, 2007). Until recently, *Xanthomonas citri* pv. *citri* TALE genes were thought to only locate on plasmids, but a chromosomal TALE has recently been reported in *Xanthomonas citri* pv. *citri* strains with a restricted host range (known as *Xanthomonas citri* pv. *citri* pathotype A^w) (Munoz Bodnar et al., 2017). In our study, we report TALEs on the chromosome of *Xanthomonas citri* pv. *citri* strains with a broad host range, including all citrus cultivars (pathotype A). Tn*3* family transposons, known to be important for the evolution of *Xanthomonas citri* pv. *citri* (Ferreira et al., 2015; Gochez et al., 2018), might be involved in the transfer of TALE from plasmid to chromosome. Deciphering the function of these TALE genes and how their chromosomal integration affects strain fitness are exciting areas for future research.

In this study, we reconstructed the genomic evolution of a lineage of the major bacterial citrus pathogen *Xanthomonas citri* pv. *citri*, from its emergence in South West Indian Ocean in the 19th century to its current endemicity. This lineage displayed a low nucleotide substitution rate, characteristic of monomorphic bacteria. In contrast to this apparent slow pace of evolution, the short-term rate of gene turnover was high. Plasmids played a key role in gene-based evolution. In particular, a copper resistance plasmid spread among the South West Indian Ocean lineage and pathogenicity-related plasmids underwent extensive evolution. By favouring intra-cellular recombination, mobile genetic elements appeared to promote genomic plasticity and migration of genetic material from plasmids to the bacterial chromosome, including the first observation of a transfer of

TALE virulence genes to the chromosome in a broad host range *Xanthomonas citri* strain. Taken together, our results highlight a far more dynamic and complex picture of ongoing adaptation in a major pathogenic crop that may have been predicted from the rate of evolution of its core genome.

To conclude, our study emphasizes the significance of plasmids as a source of gene content variation in a monomorphic bacterium. It further supports a previous study outlining the importance of plasmid-mediated horizontal gene transfer for adaptation to environmental changes in *Xanthomonas* (Halary et al., 2010). Herein, this was exemplified with the emergence and dissemination of a large conjugative plasmid conferring copper resistance in response to the massive application of copper-based pesticides used for plant disease management.

Materials and Methods

Strain collection and sequencing

The strain collection comprised 284 *Xanthomonas citri* pv. *citri* strains, of which 210 originate from the South West Indian Ocean region. Eleven strains from Martinique were also included in our sampling because previous work uncovered a likely epidemiological link between strains from Martinique and those of the South West Indian Ocean (Richard, Ravigne, et al., 2017). Additionally, the strain collection also comprised 63 strains representing the known worldwide diversity (including 17 countries, Figure 1, Supplementary Table S1). All the strains were sequenced using Illumina paired-end 2x150bp sequencing. Long-read Oxford Nanopore MinION sequencing was also performed on 13 selected strains. Copper-resistance phenotypes were determined previously (Richard, Tribot, et al., 2017). DNA sequencing and data processing are detailed in (Richard et al., 2020).

SNP detection

We used a custom bioinformatics pipeline to obtain a filtered set of SNPs from the Illumina raw reads (Supplementary Figure S8). In short, after a quality control trimming step using Trimmomatic v. 0.36 (Bolger, Lohse, & Usadel, 2014), reads were aligned against the chromosome of *Xanthomonas citri* pv. *citri* strain IAPAR 306 (GenBank accession NC_003919.1) with BWA-MEM v. 0.7.15 (Li, 2013). *Xanthomonas citri* pv. *citri* strains displayed a mean coverage of 232 (Table S1). Duplicated reads were removed using PicardTools MarkDuplicates v. 2.7. Indel realignment and SNP calling were performed using Freebayes 0.9.21-5 (Garrison & Marth, 2012). SNPs were then filtered based on allele number, coverage, phred quality, allele frequency or genomic characteristics, such as SNP density or the presence of repeated genomic regions. Recombinant regions were detected using ClonalFrameML (Didelot & Wilson, 2015) and RDP4 (Martin, Murrell, Golden, Khoosal, & Muhire, 2015) using default parameters. Recombination analysis using ClonalframeML revealed a 136kb recombinant region, comprising 233 SNPs in seven strains originating from Mali, Senegal, Bangladesh and India. Using RDP4, a 197kb (315 SNPs) recombinant region that entirely overlapped the ClonalFrameML recombinant region was detected. These 315 SNPs were thus excluded from the worldwide dataset for subsequent phylogenetic reconstruction (but retained for every South West Indian Ocean clade analysis as these strains displayed no trace of recombination).

We first tested the adequacy of several models of molecular evolution with our SNP set. According to PartitionFinder v.2.1.1 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2016) and based on Bayesian Information Criterion, the model of evolution best fitting our dataset was a General Time-Reversible substitution model of evolution with variation among sites modelled with a discrete Gamma distribution and

Invariant sites (GTR+G+I). We reconstructed a Maximum Likelihood tree of the global dataset using RAxML v.8.2.9 (Kozlov, Darriba, Flouri, Morel, & Stamatakis, 2018). The presence of the temporal signal in the dataset was tested by computing the linear regression between sample age and root-to-tip distances at every internal node of the Maximum Likelihood tree (Doizy, Prin, Cornu, Chiroleu, & Rieux, 2020). The South West Indian Ocean clade root was the deepest node for which both the linear regression was statistically significant and was, therefore, assumed to contain detectable amounts of evolutionary change, making it suitable for tip-dating inferences. For this clade, we also performed a first date-randomization test with 20 replicates using the TipDatingBeast R package (Rieux & Khatchikian, 2017) to confirm the presence of the temporal signal (Duchene, Duchene, Holmes, & Ho, 2015). Finally, we investigated whether our dataset shown confounding between temporal and genetic structures using a mantel confounding test (i.e., are closely related sequences more likely to have been sampled at similar times?), as both the root-to-tip regression and the date-randomization test can be misled in such a situation (Murray et al., 2016).

Tip-dating inference was then performed on the South West Indian Ocean subset using BEAST v1.8.4 (Drummond & Rambaut, 2007) with a GTR+G+I substitution model of evolution. We used an uncorrelated lognormal relaxed clock to account for rate variation among lineages. To minimize prior assumptions about demographic history, we first used an extended Bayesian skyline plot to integrate data over different coalescent histories. After inspecting the demographic reconstruction, an exponential growth was established as a best fit for the tree prior. Three independent chains were run for 100,000,000 steps and sampled every 10,000 steps. The first 1,000 samples (10%) were discarded as burn-in. Convergence to the stationary distribution and sufficient sampling and mixing were checked by inspecting posterior samples using Tracer v1.6 (effective sample size >200) (Rambaut, Suchard, Xie, & Drummond, 2014). After combining the three runs, a maximum clade credibility tree was obtained with TreeAnnotator v1.10.2 (Drummond & Rambaut, 2007).

Core and accessory genome assignation

Our estimation of the gene content of each strain from the South West Indian Ocean clade relied on a two-step approach. We first estimated the total South West Indian Ocean homologous set of genes with a pipeline combining *de novo* assembly, gene prediction and gene clustering (Supplementary Figure S9). Trimmed reads were mapped with BWA-MEM against six high-quality consensus chromosomes and plasmids from our previous study (Richard, Ravigne, et al., 2017). Importantly, to maximise the proportion of reads that were recruited by reference sequences, these six references included four sequence sets that corresponded to strains from the outbreak under study. Per strain, unmapped reads were then assembled using Spades v. 3.6.2 (Bankevich et al., 2012) using default parameters (assemblies details in Table S1) and genes were predicted from the resulting

contigs using prodigal v. 2.6.3 (Hyatt et al., 2010). Contigs corresponding to human or virus DNA and most probably originating from contaminants were identified using BLASTn against NCBI databases and discarded. Genes with a coverage inferior to 20 or detected as potential chimeras using the Uchime algorithm of the VSEARCH package were also discarded (Rognes, Flouri, Nichols, Quince, & Mahe, 2016). These predicted genes were added to the 29,164 from the six references mentioned above and together represented the total gene set of the South West Indian Ocean clade (35,527 genes). The total gene set was then clustered using MMseqs2 linclust using parameters -c 0.5 and --min-seq-id 0.9 (Steinegger & Soding, 2017). A total of 5,046 gene clusters was obtained. The gene content of each of the 221 strains of the South West Indian Ocean clade (on which the tip-dating inference was performed) was then defined based on the mapping of the reads of each strain on the total gene set, only considering genes with a depth of 20X over at least 60% of their length.

To validate our gene content analysis method, we first analysed three sequenced replicates of two strains, both already sequenced using the long-read Pacific Biosciences RSII technology. Besides assessing the error rate associated with our pipeline, the replicates were used to tune the mapping parameters that were used to consider a gene present in a strain. Importantly, distinct bacterial cultures and DNA extractions were performed for each of the replicates. Therefore, our estimate will confound the genuine variation resulting from gene loss during culture and the variation associated with the bioinformatics pipeline. The results of our pipeline were compared to that obtained after *de novo* assembly of the replicates using SPAdes (Bankevich et al., 2012), gene prediction using Prodigal v2.6.3 (Hyatt et al., 2010) and core genome estimation using Roary (Page et al., 2015) (Supplementary Figure S4).

Genomic location of gene clusters to the chromosome or plasmids was informed by the six high quality reference genomes. Genomic location of unassigned gene clusters was defined based on the location of genes co-occurring on the same contigs. Functions were assigned to the gene clusters according to their amino-acid homology (using a 30% identity/30% length threshold) with known Clusters of Orthologous Groups (COGs) based on a BLASTx search.

With the exception of the 13 Minion-sequenced strains (Supplementary Table S2), the nature of our sequencing data prevented us from assembling closed, circularized genomes for all the *Xanthomonas citri* pv. *citri* strain collection. However, using the gene content of each strain, we could assess the presence of genes previously identified on closed genomes. Reference circular plasmid sequences were used as the genomic reference for each plasmid's gene content and synteny: pLJ207-7.3 (GenBank accession CP018853.1, from the South West Indian Ocean strain *Xanthomonas citri* pv. *citri* LJ207-7, related to pXac47 (Martins et al., 2016)), pLH276.2 (GenBank accession CP018856.1, from the South West Indian Ocean strain *Xanthomonas citri* pv. *citri* LH276,

related to pXac64 (da Silva et al., 2002)), and the copper-resistance plasmid pCu^R (GenBank accession CP018859, from the South West Indian Ocean strain *Xanthomonas citri* pv. *citri* LH201 (Richard, Ravigne, et al., 2017)).

Subsequently, in a second BEAST analysis using the previously inferred tree topology, a discrete model was used to reconstruct the ancestral states of gene presence/absence. Three independent chains were run for 200,000,000 steps and sampled every 10,000 steps. The first 1,000 samples (10%) were discarded as burn-in. At each sampled step, we obtained a phylogenetic tree whose branches were annotated with all state changes (presence to absence or absence to presence) of each gene clusters. Gene turnover rates, gene gain rates and gene loss rates were calculated as the number of "gain and loss", "gains" and "loss" divided by the product of the number of genes of the genomic replicon considered (156 genes for plasmids and 4,571 for chromosomes) and the size of the tree in years. Turnover rates were therefore expressed in number of state change per gene and per year. The 95% HPD intervals were computed using the R HDInterval package.

Figure captions

Figure 1. Geographical location of the *Xanthomonas citri* pv. *citri* strain collection. Martinique and the islands of the South West Indian Ocean region on which this study focuses are highlighted in light blue boxes. The number of strains per location is indicated.

Figure 2. A dated phylogeny of the South West Indian Ocean clade of *Xanthomonas citri* pv. *citri*. X-axis under the phylogenetic tree represents the timescale in years (AD). Node bars represent 95% highest posterior density (HPD) for node ages estimated with tip-calibration. Tips are coloured according to sampling location. Coloured boxes indicate the plasmid content of each strain. The colour code for the boxes matches the colours for the represented circular plasmids and corresponds to observed plasmid alleles. Node numbers correspond to those in Table 1. White circle: node posterior probability >0.70; black circle: node posterior probability >0.95.

Figure 3. Representation of the gene content of the chromosomal genomic island (corresponding, in LH276, to positions 2,832,588 – 3,003,260) for 13 strains sequenced using long reads. Blocks represent genes homologous to known *Xanthomonas citri* pv. *citri* genes and are coloured according to the predicted function of their encoded proteins (see legend). Number of repeated variable di-residues of transcription-activator-like effector (TALE) genes, as estimated based on error-prone long reads, are written in the corresponding boxes.

Table 1. Inferred dates of MRCA and substitution rates of the South West Indian Ocean clade along with those of five clades of interest.

Supplementary material captions

Figure S1. (A) Maximum likelihood tree of the global dataset. Boxes are coloured according to the strains' geographic origin (see legend). Node corresponding to South West Indian Ocean clade root is annotated. (B) Map of Réunion linking the strains from Réunion to the groves (black circles) and the nursery (the blue circle) from where they were obtained. White circle: bootstrap value > 70%; black circle: bootstrap value > 95%.

Figure S2. Estimated South West Indian Ocean mean substitution rate (A) and tree root age (B) obtained from 20 BEAST analysis ran with randomized sample dates (black) and from the real dataset (red). Black circles represent mean values and intervals represent associated 95% HPD. (C) and (D) show similar plots obtained after date randomion among clusters of strains isolated at similar dates. (E) Root-to-tip regression between root-to-tip distance and tip dates.

Figure S3. Bayesian phylogeny (left) and the Maximum Likelihood phylogeny (right) of the South West Indian Ocean strains. The correspondence of strains is displayed using the links drawn in-between the trees. Tips are coloured according to sampling location, using the same colour code as in Figure 2.

Figure S4. Venn diagram representing the repeatability of the gene content prediction for replicates (noted A, B and C) of LH201 (left) and LE50 (right). The in-house pipeline (hybrid mapping – *de novo* pipeline) is compared to a full *de novo* pipeline where, after contigs assembly using SPAdes and gene prediction using, Roary is employed to infer the core genome. Whereas in the least conservative version of the pipeline, default parameters are used for the different programs, in the most conservative version, genes are not allowed to run off contig edges (Prodigal "-c" parameter) and we disabled the splitting of paralog groups (Roary "-s" option). Due to intrinsic differences between the tested approaches, note that the total number of predicted genes vary greatly among methods.

Figure S5. Time-calibrated phylogeny of the South West Indian Ocean clade. The 221 rows by 699 columns matrix on the right of the tree comprises one column for each of the 699 genes that varied in presence/absence among the 221 South West Indian Ocean strains. A cell is coloured if the corresponding strain (rows) carries the corresponding gene cluster (columns). Gene clusters are separated according to their genetic compartment of origin: plasmid (green), chromosome (red) or undefined (purple). Tips with asterisks indicate strains for

which long-read sequencing was also performed. White circle: node posterior probability >0.70; black circle: node posterior probability >0.95.

Figure S6. Bayesian phylogeny of the South West Indian Ocean which branches are coloured according to their mean chromosomal gene turnover rate (A) and plasmid gene turnover rate (B). Boxplots of chromosomal (C) and plasmid (D) gene turnover rates are further represented as a function of the mean age of the branches of the phylogeny, for five branch age intervals.

Figure S7. Bayesian phylogeny of the South West Indian Ocean strains based on their chromosomal SNPs (left) and phylogeny of the copper-resistance plasmid present in some strains (right). Dotted lines join each strain with its plasmid. Blue circles at the tip of the branches of the phylogeny represent strains isolated in a Réunion nursery.

Figure S8. Schematics of the bioinformatics pipeline used for SNP inference.

Figure S9. Schematics of the bioinformatics pipeline used to estimate the total SWIO homologous set of genes.

Table S1. External Excel document. Characteristics of all the sequenced bacterial strains of this study and metrics related to sequencing and steps of the bioinformatics pipeline applied on the dataset. Copper phenotype of the strains was abbreviated as S for copper-sensitivity, Rlab for copper-resistance due to the *CopLAB* system and Rabcd for copper-resistance due to the *CopABCD* system.

Table S2. External Excel document. Metrics of hybrid *de novo* assemblies (Nanopore MinION and Illumina) of 13 *Xanthomonas citri* pv. *citri* strains.

Data availability

The sequencing data generated in this study were published in the NCBI GenBank repository under accession numbers listed in Supplementary Table S1, Additional File 1.

Acknowledgements

We would like to express our thanks to the plant protection platform (3P, IBISA) and to S. Javegny, K. Boyer and J. Hascoat for their helpful contribution. The European regional development fund (ERDF contract GURDT I2016-1731-0006632), the Réunion region, the French government, the French agropolis foundation (Labex Agro – Montpellier, E-SPACE project number 1504-004), ANSES and CIRAD provided financial support. We would like to thank INRAPE (the Union of the Comoros), FAREI (Mauritius) and NBA (Seychelles) for providing us with diseased citrus material. This work was supported by the CIRAD - UMR AGAP HPC data center of the south green bioinformatics platform (http://www.southgreen.fr/).

Funding

The European regional development fund (ERDF contract GURDT I2016-1731-0006632), the Réunion region, the French government, the French agropolis foundation (Labex Agro – Montpellier, E-SPACE project number 1504-004), ANSES and CIRAD provided financial support.

Author contributions

D.R., O.P. and P.L. designed and conceived the study; C.B. processed the samples in the wet-lab; D.R. performed computational analyses with inputs from P.L. and A.R.; D.R., P.L. and O.P. wrote the manuscript with inputs from all co-authors.

References

Achtman, M. (2012). Insights from genomic comparisons of genetically monomorphic bacterial pathogens. *Philos Trans R Soc Lond B Biol Sci, 367*(1590), 860-867. doi:10.1098/rstb.2011.0303

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol Evol, 19*(10), 535-544. doi:10.1016/j.tree.2004.07.021

Andreani, N. A., Hesse, E., & Vos, M. (2017). Prokaryote genome fluidity is dependent on effective population size. *The ISME journal, 11*(7), 1719-1721. doi:10.1038/ismej.2017.36

Aubert, B. (2014). Vergers de la Réunion et de l'Océan Indien. In M. CIRAD, France (Ed.), *Hommes et fruits en pays du Sud* (pp. 111–165).

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., . . . Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol, 19*(5), 455-477. doi:10.1089/cmb.2012.0021

Barash, I., & Manulis-Sasson, S. (2009). Recent evolution of bacterial pathogens: the gall-forming *Pantoea agglomerans* case. *Annual Review of Phytopathology, 47*, 133-152. doi:10.1146/annurev-phyto-080508-081803

Behlau, F., Hong, J. C., Jones, J. B., & Graham, J. H. (2013). Evidence for acquisition of copper resistance genes from different sources in citrus-associated xanthomonads. *Phytopathology, 103*(5), 409-418. doi:10.1094/PHYTO-06-12-0134-R

Bennett, P. M. (2004). Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement. *Methods Mol Biol, 266*, 71-113. doi:10.1385/1-59259-763-7:071

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Bolotin, E., & Hershberg, R. (2017). Horizontally acquired genes are often shared between closely related bacterial species. *Frontiers in Microbiology, 8*, 1536. doi:10.3389/fmicb.2017.01536

Campbell, P. C. (1923). *Chinese coolie emigration to countries within the British empire* (U. K. S. Westminster Ed.).

Carter, M. T., K. (2002). *Coolitude: an anthology of the Indian labour diaspora* (U. A. P. London Ed.).

Chen, N. W. G., Serres-Giardi, L., Ruh, M., Briand, M., Bonneau, S., Darrasse, A., . . . Jacques, M. A. (2018). Horizontal gene transfer plays a major role in the pathological convergence of *Xanthomonas* lineages on common bean. *Bmc Genomics, 19*(1), 606. doi:10.1186/s12864-018-4975-4

Croucher, N. J., Coupland, P. G., Stevenson, A. E., Callendrello, A., Bentley, S. D., & Hanage, W. P. (2014). Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nature Communications, 5*(1), 5471. doi:10.1038/ncomms6471

da Silva, A. C., Ferro, J. A., Reinach, F. C., Farah, C. S., Furlan, L. R., Quaggio, R. B., . . . Kitajima, J. P. (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature, 417*(6887), 459-463. doi:10.1038/417459a

Didelot, X., & Wilson, D. J. (2015). ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol, 11*(2), e1004041. doi:10.1371/journal.pcbi.1004041

Doizy, A., Prin, A., Cornu, G., Chiroleu, F., & Rieux, A. (2020). Phylostems: a new graphical tool to investigate temporal signal of heterochronous sequences at various evolutionary scales. 2020.2010.2019.346429. doi:10.1101/2020.10.19.346429 %J bioRxiv

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol, 7*(1), 214. doi:10.1186/1471-2148-7-214

Du Pont de Nemours, P. S. (1797). *Oeuvres complètes de P. Poivre, intendant des isles de France et de Bourbon, correspondant de l'académie des sciences, etc* (F. F. Paris Ed.).

Duchene, S., Duchene, D., Holmes, E. C., & Ho, S. Y. (2015). The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. *Molecular Biology and Evolution, 32*(7), 1895-1906. doi:10.1093/molbev/msv056

Duchene, S., Holt, K. E., Weill, F. X., Le Hello, S., Hawkey, J., Edwards, D. J., . . . Holmes, E. C. (2016). Genome-scale rates of evolutionary change in bacteria. *Microb Genom, 2*(11), e000094. doi:10.1099/mgen.0.000094

El Yacoubi, B., Brunings, A. M., Yuan, Q., Shankar, S., & Gabriel, D. W. (2007). In planta horizontal transfer of a major pathogenicity effector gene. *Applied and Environmental Microbiology, 73*(5), 1612-1621. doi:10.1128/aem.00261-06

Engering, A., Hogerwerf, L., & Slingenbergh, J. (2013). Pathogen–host–environment interplay and disease emergence. *Emerging Microbes & Infections, 2*(1), 1-7. doi:10.1038/emi.2013.5

Estoup, A., & Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology, 19*(19), 4113-4130. doi:10.1111/j.1365-294X.2010.04773.x

Ferreira, R. M., de Oliveira, A. C., Moreira, L. M., Belasque, J., Jr., Gourbeyre, E., Siguier, P., . . . Varani, A. M. (2015). A TALE of transposition: Tn3-like transposons play a major role in the spread of pathogenicity determinants of *Xanthomonas citri* and other xanthomonads. *MBio, 6*(1), e02505-02514. doi:10.1128/mBio.02505-14

Fogg, P. C., Colloms, S., Rosser, S., Stark, M., & Smith, M. C. (2014). New applications for phage integrases. *Journal of Molecular Biology, 426*(15), 2703-2716. doi:10.1016/j.jmb.2014.05.014

Gardan, L., Brault, T., & Germain, E. (1993). *Copper resistance of Xanthomonas campestris pv juglandis in French walnut orchards and its association with conjugative plasmids.* Paper presented at the International Walnut Meeting 311.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Gochez, A. M., Huguet-Tapia, J. C., Minsavage, G. V., Shantaraj, D., Jalan, N., Strauss, A., . . . Potnis, N. (2018). Pacbio sequencing of copper-tolerant *Xanthomonas citri* reveals presence of a chimeric plasmid structure and provides insights into reassortment and shuffling of transcription activator-like effectors among *X. citri* strains. *Bmc Genomics, 19*(1), 16. doi:10.1186/s12864-017-4408-9

Gogarten, J. P., & Townsend, J. P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol, 3*(9), 679-687. doi:10.1038/nrmicro1204

Gordon, J. L., Lefeuvre, P., Escalon, A., Barbe, V., Cruveiller, S., Gagnevin, L., & Pruvost, O. (2015). Comparative genomics of 43 strains of *Xanthomonas citri* pv. *citri* reveals the evolutionary events giving rise to pathotypes with different host ranges. *Bmc Genomics, 16*, 1098. doi:10.1186/s12864-015-2310-x

Govindin, S.-S. (1994). *Les engagés indiens: Ile de la Réunion, XIXe siècle* (A. Éditions Ed.).

Graham, J. H., Gottwald, T. R., Cubero, J., & Achor, D. S. (2004). *Xanthomonas axonopodis* pv. *citri*: factors affecting successful eradication of citrus canker. *Mol Plant Pathol, 5*(1), 1-15. doi:10.1046/j.1364-3703.2004.00197.x

Graña-Miraglia, L., Lozano, L. F., Velázquez, C., Volkow-Fernández, P., Pérez-Oseguera, Á., Cevallos, M. A., & Castillo-Ramírez, S. (2017). Rapid gene turnover as a significant source of genetic variation in a recently seeded population of a healthcare-associated pathogen. *8*(1817). doi:10.3389/fmicb.2017.01817

Guillemaud, T., Ciosi, M., Lombaert, É., & Estoup, A. (2011). Biological invasions in agricultural settings: Insights from evolutionary biology and population genetics. *Comptes Rendus Biologies, 334*(3), 237-246. doi:https://doi.org/10.1016/j.crvi.2010.12.008

Haegeman, B., & Weitz, J. S. (2012). A neutral theory of genome evolution and the frequency distribution of genes. *Bmc Genomics, 13*(1), 196. doi:10.1186/1471-2164-13-196

Halary, S., Leigh, J. W., Cheaib, B., Lopez, P., & Bapteste, E. (2010). Network analyses structure genetic diversity in independent genetic worlds. *Proc Natl Acad Sci U S A, 107*(1), 127-132. doi:10.1073/pnas.0908978107

Hall, J. P. J., Brockhurst, M. A., & Harrison, E. (2017). Sampling the mobile gene pool: innovation via horizontal gene transfer in bacteria. *Philos Trans R Soc Lond B Biol Sci, 372*(1735), 20160424. doi:10.1098/rstb.2016.0424

Hao, W., & Golding, G. B. (2006). The fate of laterally transferred genes: life in the fast lane to adaptation or death. *Genome Res, 16*(5), 636-643. doi:10.1101/gr.4746406

Hulme, P. E. (2009). Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology, 46*(1), 10-18. doi:doi:10.1111/j.1365-2664.2008.01600.x

Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *Bmc Bioinformatics, 11*, 119. doi:10.1186/1471-2105-11-119

Hyun, J.-W., Kim, H.-J., Yi, P.-H., Hwang, R.-Y., & Park, E.-W. (2012). Mode of action of streptomycin resistance in the citrus canker pathogen (*Xanthomonas smithii* subsp. *citri*) in Jeju Island. *The Plant Pathology Journal, 28*(2), 207-211. doi:10.5423/PPJ.2012.28.2.207

Iranzo, J., Wolf, Y. I., Koonin, E. V., & Sela, I. (2019). Gene gain and loss push prokaryotes beyond the homologous recombination barrier and accelerate genome sequence divergence. *Nature Communications, 10*(1), 5376. doi:10.1038/s41467-019-13429-2

Jacques, M. A., Arlat, M., Boulanger, A., Boureau, T., Carrere, S., Cesbron, S., . . . Verniere, C. (2016). Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annual Review of Phytopathology, 54*, 163-187. doi:10.1146/annurev-phyto-080615-100147

Kado, C. I. (2015). Historical events that spawned the field of plasmid biology. In *Plasmids: Biology and Impact in Biotechnology and Discovery* (pp. 3-11): American Society of Microbiology.

Kozlov, A., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2018). RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *bioRxiv*, 447110. doi:10.1093/bioinformatics/btz305

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution, 34*(3), 772-773. doi:10.1093/molbev/msw260

Leduc, A., Traore, Y. N., Boyer, K., Magne, M., Grygiel, P., Juhasz, C. C., . . . Pruvost, O. (2015). Bridgehead invasion of a monomorphic plant pathogenic bacterium: *Xanthomonas citri* pv. *citri*, an emerging citrus pathogen in Mali and Burkina Faso. *Environmental Microbiology, 17*(11), 4429-4442. doi:10.1111/1462-2920.12876

Lerat, E., Daubin, V., Ochman, H., & Moran, N. A. (2005). Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol, 3*(5), e130. doi:10.1371/journal.pbio.0030130

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Lima, W. C., Paquola, A. C. M., Varani, A. M., Van Sluys, M.-A., & Menck, C. F. M. (2008). Laterally transferred genomic islands in Xanthomonadales related to pathogenicity and primary metabolism. *Fems Microbiology Letters, 281*(1), 87-97. doi:10.1111/j.1574-6968.2008.01083.x

Lobkovsky, A. E., Wolf, Y. I., & Koonin, E. V. (2013). Gene frequency distributions reject a neutral model of genome evolution. *Genome Biol Evol, 5*(1), 233-242. doi:10.1093/gbe/evt002

Marri, P. R., Hao, W., & Golding, G. B. (2006). Gene gain and gene loss in *Streptococcus*: is it driven by habitat? *Molecular Biology and Evolution, 23*(12), 2379-2391. doi:10.1093/molbev/msl115

Marri, P. R., Hao, W., & Golding, G. B. (2007). The role of laterally transferred genes in adaptive evolution. *BMC Evol Biol, 7 Suppl 1*(S1), S8. doi:10.1186/1471-2148-7-S1-S8

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol, 1*(1), vev003. doi:10.1093/ve/vev003

Martins, P. M., Machado, M. A., Silva, N. V., Takita, M. A., & de Souza, A. A. (2016). Type II toxin-antitoxin distribution and adaptive aspects on *Xanthomonas* genomes: focus on *Xanthomonas citri*. *Front Microbiol, 7*, 652. doi:10.3389/fmicb.2016.00652

McDonald, B. A., & Stukenbrock, E. H. (2016). Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philos Trans R Soc Lond B Biol Sci, 371*(1709), 20160026. doi:10.1098/rstb.2016.0026

McManus, P. S., Stockwell, V. O., Sundin, G. W., & Jones, A. L. (2002). Antibiotic use in plant agriculture. *Annual Review of Phytopathology, 40*(1), 443-465. doi:10.1146/annurev.phyto.40.120301.093927

Minsavage, G. V., Canteros, B. I., & Stall, R. E. (1990). Plasmid-mediated resistance to streptomycin in *Xanthomonas campestris* pv. *vesicatoria*. *Phytopathology, 80*(8), 719-723.

Monteil, C. L., Yahara, K., Studholme, D. J., Mageiros, L., Meric, G., Swingle, B., . . . Sheppard, S. K. (2016). Population-genomic insights into emergence, crop adaptation and dissemination of *Pseudomonas syringae* pathogens. *Microb Genom, 2*(10), e000089. doi:10.1099/mgen.0.000089

Munoz Bodnar, A., Santillana, G., Mavrodieva, V., Liu, Z., Nakhla, M., & Gabriel, D. W. (2017). Complete genome sequences of three *Xanthomonas citri* strains from Texas. *Genome Announc, 5*(28), e00609-00617. doi:10.1128/genomeA.00609-17

Murray, G. G., Wang, F., Harrison, E. M., Paterson, G. K., Mather, A. E., Harris, S. R., . . . Welch, J. J. (2016). The effect of genetic structure on molecular dating and tests for temporal signal. *Methods Ecol Evol, 7*(1), 80-89. doi:10.1111/2041-210X.12466

Niu, X. N., Wei, Z. Q., Zou, H. F., Xie, G. G., Wu, F., Li, K. J., . . . He, Y. Q. (2015). Complete sequence and detailed analysis of the first indigenous plasmid from *Xanthomonas oryzae* pv. *oryzicola*. *Bmc Microbiology, 15*, 233. doi:10.1186/s12866-015-0562-x

Norman, A., Hansen, L. H., & Sorensen, S. J. (2009). Conjugative plasmids: vessels of the communal gene pool. *Philos Trans R Soc Lond B Biol Sci, 364*(1527), 2275-2289. doi:10.1098/rstb.2009.0037

Nowell, R. W., Green, S., Laue, B. E., & Sharp, P. M. (2014). The extent of genome flux and its role in the differentiation of bacterial lineages. *Genome Biol Evol, 6*(6), 1514-1529. doi:10.1093/gbe/evu123

Osborn, A. M., & Böltner, D. (2002). When phage, plasmids, and transposons collide: genomic islands, and conjugative-and mobilizable-transposons as a mosaic continuum. *Plasmid, 48*(3), 202-212. doi:10.1016/s0147-619x(02)00117-8

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., . . . Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics, 31*(22), 3691-3693. doi:10.1093/bioinformatics/btv421

Pilla, G., & Tang, C. M. (2018). Going around in circles: virulence plasmids in enteric pathogens. *Nat Rev Microbiol, 16*(8), 484-495. doi:10.1038/s41579-018-0031-2

Pruvost, O., Magne, M., Boyer, K., Leduc, A., Tourterel, C., Drevet, C., . . . Verniere, C. (2014). A MLVA genotyping scheme for global surveillance of the citrus pathogen *Xanthomonas citri* pv. *citri* suggests a worldwide geographical expansion of a single genetic lineage. *Plos One, 9*(6), e98129. doi:10.1371/journal.pone.0098129

Rambaut, A., Suchard, M., Xie, D., & Drummond, A. (2014). Tracer, version 1.6, MCMC trace analysis package [Internet]. In.

Richard, D., Ravigne, V., Rieux, A., Facon, B., Boyer, C., Boyer, K., . . . Lefeuvre, P. (2017). Adaptation of genetically monomorphic bacteria: evolution of copper resistance through multiple horizontal gene transfers of complex and versatile mobile genetic elements. *Mol Ecol, 26*(7), 2131-2149. doi:10.1111/mec.14007

Richard, D., Rieux, A., Lefeuvre, P., Hamza, A., Lobin, K. K., Naiken, M., . . . Pruvost, O. (2020). Complete genome sequences of 284 *Xanthomonas citri* pv. *citri* strains causing Asiatic citrus canker. *in press*.

Richard, D., Tribot, N., Boyer, C., Terville, M., Boyer, K., Javegny, S., . . . Vernière, C. (2017). First report of copper-resistant *Xanthomonas citri* pv. *citri* pathotype A causing asiatic citrus canker in Réunion, France. *Plant Disease, 101*(3). doi:10.1094/pdis-09-16-1387-pdn

Rieux, A., & Khatchikian, C. E. (2017). TipDatingBeast: An R package to assist the implementation of phylogenetic tip-dating tests using BEAST. *Molecular Ecology Resources, 17*(4), 608-613. doi:10.1111/1755-0998.12603

Rocha, E. P. (2008). Evolutionary patterns in prokaryotic genomes. *Curr Opin Microbiol, 11*(5), 454-460. doi:10.1016/j.mib.2008.09.007

Roeschlin, R. A., Uviedo, F., Garcia, L., Molina, M. C., Favaro, M. A., Chiesa, M. A., . . . Marano, M. R. (2019). PthA4(AT) , a 7.5-repeats transcription activator-like (TAL) effector from *Xanthomonas citri* ssp. *citri*, triggers citrus canker resistance. *Mol Plant Pathol, 20*(10), 1394-1407. doi:10.1111/mpp.12844

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahe, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ, 4*, e2584. doi:10.7717/peerj.2584

Ruh, M., Briand, M., Bonneau, S., Jacques, M.-A., & Chen, N. W. (2017). *Xanthomonas* adaptation to common bean is associated with horizontal transfers of genes encoding TAL effectors. *Bmc Genomics, 18*(1), 670. doi:10.1186/s12864-017-4087-6

Shiotani, H., Fujikawa, T., Ishihara, H., Tsuyumu, S., & Ozaki, K. (2007). A pthA homolog from *Xanthomonas axonopodis* pv. *citri* responsible for host-specific suppression of virulence. *Journal of Bacteriology, 189*(8), 3271-3279. doi:10.1128/JB.01790-06

Steinegger, M., & Soding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol, 35*(11), 1026-1028. doi:10.1038/nbt.3988

Streubel, J., Blucher, C., Landgraf, A., & Boch, J. (2012). TAL effector RVD specificities and efficiencies. *Nat Biotechnol, 30*(7), 593-595. doi:10.1038/nbt.2304

Swarup, S., De Feyter, R., Brlansky, R. H., & Gabriel, D. W. (1991). A pathogenicity locus from *Xanthomonas citri* enables strains from several pathovars of *X. campestris* to elicit cankerlike lesions on citrus. *Phytopathology, 81*(7), 802-809.

Thynne, E., McDonald, M. C., & Solomon, P. S. (2015). Phytopathogen emergence in the genomics era. *Trends Plant Sci, 20*(4), 246-255. doi:10.1016/j.tplants.2015.01.009

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., . . . Denamur, E. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet, 5*(1), e1000344. doi:10.1371/journal.pgen.1000344

Vinatzer, B. A., Monteil, C. L., & Clarke, C. R. (2014). Harnessing population genomics to understand how bacterial pathogens emerge, adapt to crop hosts, and disseminate. *Annual Review of Phytopathology, 52*, 19-43. doi:10.1146/annurev-phyto-102313-045907

Vos, M., Hesselman, M. C., Te Beek, T. A., van Passel, M. W. J., & Eyre-Walker, A. (2015). Rates of lateral gene transfer in prokaryotes: high but why? *Trends in Microbiology, 23*(10), 598-605. doi:10.1016/j.tim.2015.07.006

Weber, B. S., Ly, P. M., Irwin, J. N., Pukatzki, S., & Feldman, M. F. (2015). A multidrug resistance plasmid contains the molecular switch for type VI secretion in *Acinetobacter baumannii*. *Proc Natl Acad Sci U S A, 112*(30), 9442-9447. doi:10.1073/pnas.1502966112

Wu, G. A., Terol, J., Ibanez, V., Lopez-Garcia, A., Perez-Roman, E., Borreda, C., . . . Talon, M. (2018). Genomics of the origin and evolution of *Citrus*. *Nature, 554*(7692), 311-316. doi:10.1038/nature25447

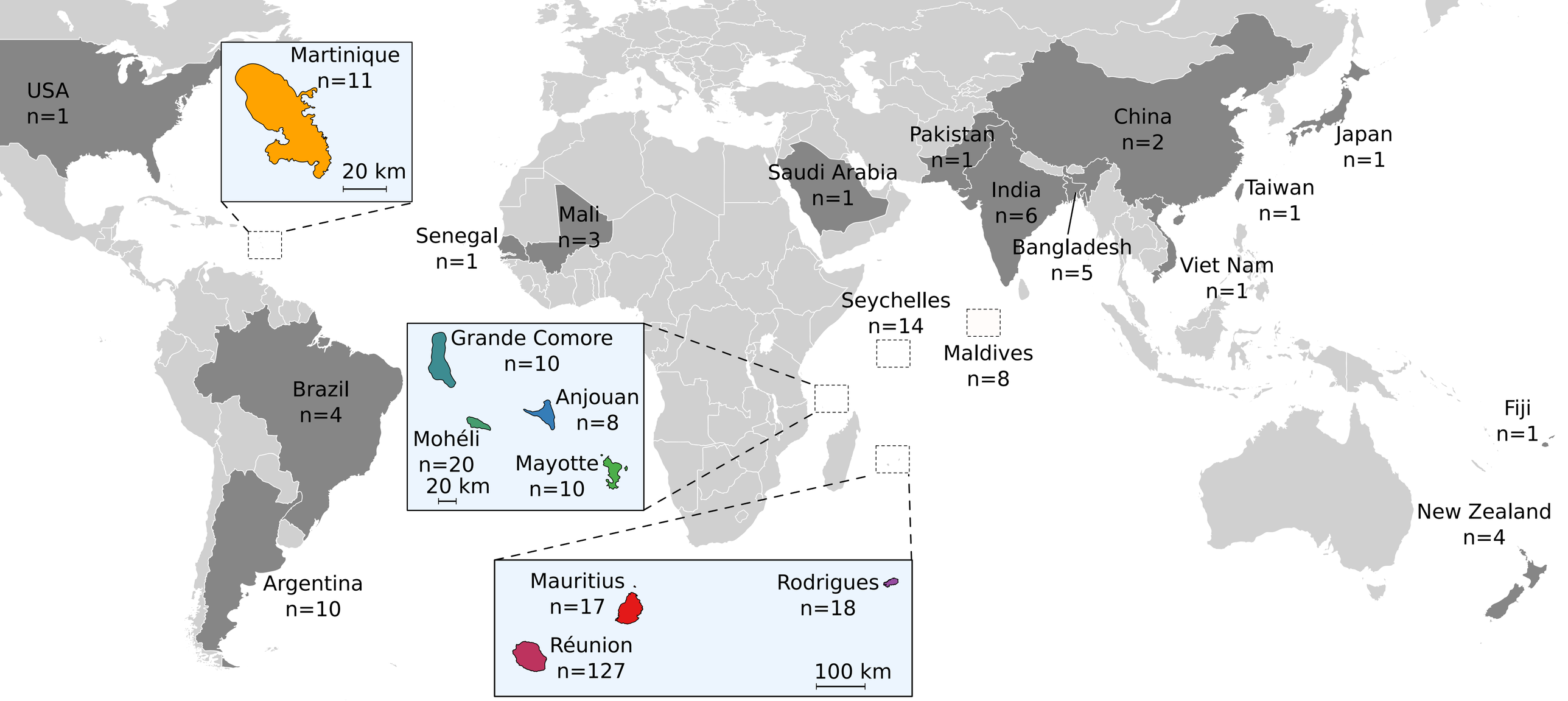
Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., . . . Krause, J. (2013). The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *Elife, 2*, e00731. doi:10.7554/eLife.00731

Zhou, Z., Lundstrøm, I., Tran-Dien, A., Duchêne, S., Alikhan, N.-F., Sergeant, M. J., . . . Stenøien, H. K. (2018). Pan-genome analysis of ancient and modern *Salmonella enterica* demonstrates genomic stability of the invasive Para C Lineage for millennia. *Current Biology, 28*(15), 2420-2428. e2410. doi:10.1016/j.cub.2018.05.058

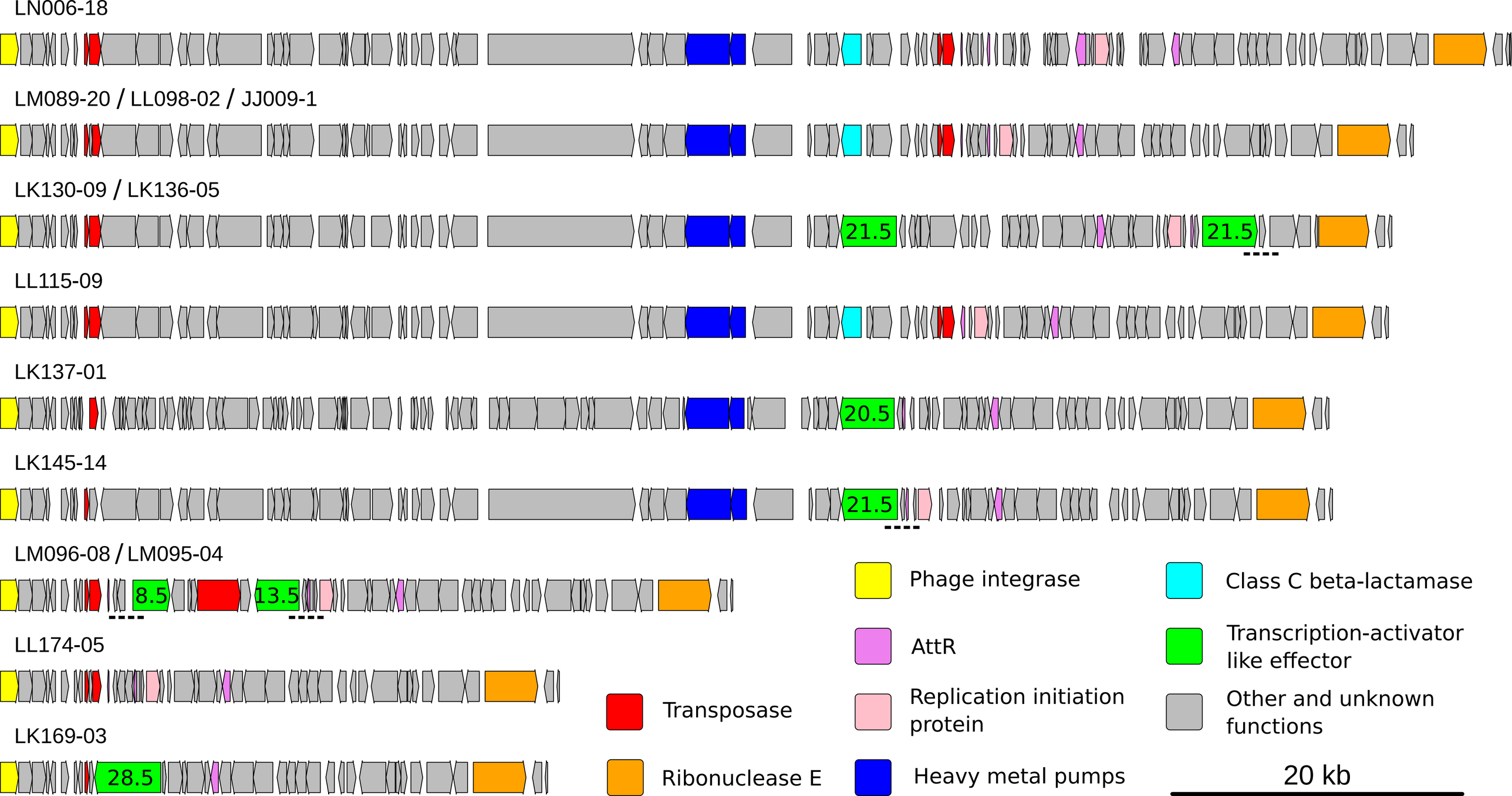
Zhou, Z., McCann, A., Litrup, E., Murphy, R., Cormican, M., Fanning, S., . . . Achtman, M. (2013). Neutral genomic microevolution of a recently emerged pathogen, *Salmonella enterica* serovar Agona. *PLoS Genet, 9*(4), e1003471. doi:10.1371/journal.pgen.1003471

Table 1. Inferred dates of MRCA and substitution rates of the SWIO clade along with those of five clades of interest.

Node number	Inferred node date (HPD95%)	No. of strains	No. of variable genes	No. of variable SNPs	Subst. rate SNP (HPD95%)	No. of core-genome genes	No. of pan-genome genes	Plasmid gene turnover rate (HPD95%)	Chromosome gene turnover rate (HPD95%)
1	1818 (1762-1868)	221	699	3403	8.4E-8 (6.9E-8 - 1.0E-7)	4347	5046	2.2E-3 (1.7E-3 - 3.4E-3)	8.8E-5 (8.1E-5 - 9.7E-5)
2	1906 (1882-1927)	48	205	812	8.8E-8 (4.6E-8 - 1.3E-7)	4546	4751	5.5E-4 (4.0E-4 - 9.6E-4)	5.3E-5 (4.8E-5 - 6.1E-5)
3	1931 (1914-1947)	19	222	363	9.0E-8 (5.5E-8 - 1.3E-7)	4531	4753	1.2E-3 (1.0E-3 - 1.6E-3)	1.1E-4 (9.9E-5 -1.2E-4)
4	1949 (1934-1963)	12	143	134	8.5E-8 (5.4E-8 - 1.2E-7)	4611	4754	2.3E-3 (8.3E-4 - 2.4E-3)	1.3E-4 (1.1E-4 - 1.6E-4)
5	1959 (1946-1971)	27	147	288	9.4E-8 (5.5E-8 - 1.4E-7)	4593	4740	8.2E-4 (5.2E-4 - 1.3E-3)	6.4E-5 (5.7E-5 - 7.5E-5)
6	1965 (1956-1973)	62	358	318	1.0E-7 (6.5E-8 - 1.4E-7)	4599	4957	5.9E-3 (5.1E-3 - 8.4E-3)	1.2E-4 (1.1E-4 - 1.3E-4)


Martinique
n=11
20 km
USA
n=1
Senegal
n=1
Mali
n=3
Saudi Arabia
n=1
Pakistan
n=1
China
n=2
Japan
n=1
India
n=6
Taiwan
n=1
Bangladesh
n=5
Viet Nam
n=1
Seychelles
n=14
Maldives
n=8
Grande Comore
n=10
Anjouan
n=8
Mohéli
n=20
Mayotte
n=10
20 km
Brazil
n=4
Fiji
n=1
New Zealand
n=4
Argentina
n=10
Mauritius
n=17
Rodrigues
n=18
Réunion
n=127
100 km

Geographic origin
pCuR
pLJ207-7.3
pLH276.2
pXac39
Plasmid alleles
pCuR
pLJ207-7.3
pLH276.2
pXac39
1
2
3
4
5
6
Martinique
Mauritius
Réunion
Rodrigues
Mayotte
Comoros - Moheli
Comoros - Anjouan
Comoros - Grande Comore
1767
1817
1867
1917
1967
2017


LN006-18
LM089-20 / LL098-02 / JJ009-1
LK130-09 / LK136-05
21.5
21.5
LL115-09
LK137-01
20.5
LK145-14
21.5
LM096-08 / LM095-04
8.5
13.5
LL174-05
LK169-03
28.5
Phage integrase
Class C beta-lactamase
AttR
Transcription-activator like effector
Transposase
Replication initiation protein
Other and unknown functions
Ribonuclease E
Heavy metal pumps
20 kb