



## Horizontal gene transfer plays a major role in the pathological convergence of *Xanthomonas* lineages on common bean

Nicolas Chen, Laurana Serres-Giardi, Mylène Ruh, Martial Briand, Sophie Bonneau, Armelle Darrasse, Valérie Barbe, Lionel Gagnevin, Ralf Koebnik, Marie Agnes Jacques

### ► To cite this version:

Nicolas Chen, Laurana Serres-Giardi, Mylène Ruh, Martial Briand, Sophie Bonneau, et al.. Horizontal gene transfer plays a major role in the pathological convergence of *Xanthomonas* lineages on common bean. BMC Genomics, 2018, 19 (1), 20 p. 10.1186/s12864-018-4975-4 . hal-02627510

**HAL Id: hal-02627510**

**<https://hal.inrae.fr/hal-02627510>**

Submitted on 26 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE

Open Access



# Horizontal gene transfer plays a major role in the pathological convergence of *Xanthomonas* lineages on common bean

Nicolas W. G. Chen<sup>1†</sup>, Laurana Serres-Giardi<sup>1†</sup>, Mylène Ruh<sup>1</sup>, Martial Briand<sup>1</sup>, Sophie Bonneau<sup>1</sup>, Armelle Darrasse<sup>1</sup>, Valérie Barbe<sup>2</sup>, Lionel Gagnevin<sup>3,4</sup>, Ralf Koebnik<sup>4</sup> and Marie-Agnès Jacques<sup>1\*</sup>

## Abstract

**Background:** Host specialization is a hallmark of numerous plant pathogens including bacteria, fungi, oomycetes and viruses. Yet, the molecular and evolutionary bases of host specificity are poorly understood. In some cases, pathological convergence is observed for individuals belonging to distant phylogenetic clades. This is the case for *Xanthomonas* strains responsible for common bacterial blight of bean, spread across four genetic lineages. All the strains from these four lineages converged for pathogenicity on common bean, implying possible gene convergences and/or sharing of a common arsenal of genes conferring the ability to infect common bean.

**Results:** To search for genes involved in common bean specificity, we used a combination of whole-genome analyses without a priori, including a genome scan based on *k*-mer search. Analysis of 72 genomes from a collection of *Xanthomonas* pathovars unveiled 115 genes bearing DNA sequences specific to strains responsible for common bacterial blight, including 20 genes located on a plasmid. Of these 115 genes, 88 were involved in successive events of horizontal gene transfers among the four genetic lineages, and 44 contained nonsynonymous polymorphisms unique to the causal agents of common bacterial blight.

**Conclusions:** Our study revealed that host specificity of common bacterial blight agents is associated with a combination of horizontal transfers of genes, and highlights the role of plasmids in these horizontal transfers.

**Keywords:** *Xanthomonas*, Common bean, TAL effectors, Host adaptation, Horizontal gene transfer

## Background

In nature, most pathogens are generalists, meaning that they are able to infect multiple hosts, while other pathogens are specialists, meaning that they are highly adapted to a single or few host species [1]. For plant pathogens, adaptation to a specific host plant is a complex process possibly involving diverse molecular determinants and leading to host specificity [2, 3]. Understanding the molecular basis of host specificity can provide new insights into the evolution and ecology of specialist pathogens, and their potential to shift species and to infect new hosts. Bacteria from the genus *Xanthomonas* infect at least 392

plant species including important crops and ornamentals [4]. Yet, each individual strain is able to infect only one or few plant species. Strains able to cause the same symptoms on the same host range are grouped into pathovars [5]. Although our understanding of the molecular basis of host specificity is still limited, chemotactic sensors, adhesins and type III effectors emerge as key determinants for shaping host specificity in *Xanthomonas* [6–8]. Chemotactic sensors enable the bacteria to detect attractant or repellent molecules and trigger flagellar motility towards entry sites of the host plant, while adhesins allow the attachment on the host plant surface and biofilm formation, and type III effectors are delivered into the plant cells where they can have different functions including providing pathogen-associated molecular pattern triggered immunity (PTI).

\* Correspondence: [marie-agnes.jacques@inra.fr](mailto:marie-agnes.jacques@inra.fr)

†Nicolas W. G. Chen and Laurana Serres-Giardi contributed equally to this work.

<sup>1</sup>IRHS, INRA, AGROCAMPUS OUEST, Université d'Angers, SFR4207 QUASAV, 42, rue Georges Morel, 49071 Beaucouzé, France

Full list of author information is available at the end of the article



The *Xanthomonas axonopodis* species complex sensu Vauterin [9, 10] groups more than 30 pathovars infecting a wide range of plants including economically important crops and ornamentals, such as *Citrus*, *Anthurium* and *Dieffenbachia* species, as well as pepper, cassava, cotton, mango, soybean, and common bean. Based on repetitive-sequence-based Polymerase Chain Reaction (rep-PCR) fingerprints, *X. axonopodis* has been subdivided into six subclusters named 9.1 to 9.6 [11]. More recently, this species complex has been split into the four species *X. citri*, *X. euvesicatoria*, *X. phaseoli* and *X. axonopodis* [12, 13]. Common bacterial blight of bean (CBB) is the most devastating bacterial disease infecting common bean (*Phaseolus vulgaris* L.). CBB occurs everywhere where common bean is cultivated and may cause up to 75% yield loss in the most severe cases [14, 15]. *Xanthomonas* strains responsible for CBB are distributed across four different genetic lineages [16]. The fuscous lineage (fuscans) and the non-fuscous lineages 2 (NF2) and 3 (NF3) belong to *X. citri* pv. *fuscans* while the non-fuscous lineage 1 (NF1) belongs to *X. phaseoli* pv. *phaseoli* [9, 11, 12]. Pathological convergence between the NF1 and fuscans lineages is associated with horizontal gene transfers (HGT) involving dozens of genes [17]. Horizontal transfer of genes encoding Transcription Activator-Like (TAL) type III effectors was also observed between the four lineages of CBB agents [18]. In particular, all strains from the four genetic lineages display an allele of the *tal23A* gene, suggesting that this gene is important for *Xanthomonas* adaptation to common bean.

In order to search for *Xanthomonas* genes putatively involved in the adaptation leading to common bean specificity in *Xanthomonas*, we have generated the whole genome sequences of 17 *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* strains. A combination of approaches including a comparison between the phylogeny of genes and the phylogeny of organisms, a parsimony approach to infer gene gains and losses, and a genome-wide search for specific *k*-mers, was used to search for genes presenting common characteristics unique to strains belonging to the four bean-pathogenic lineages of *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli*.

## Results

### Genome sequencing and phylogeny

In order to obtain genomic data representative of the diversity of *Xanthomonas* strains responsible for CBB, we produced whole genome sequences for 17 strains from the four genetic lineages of *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* that affect beans. In addition, we sequenced two strains of *X. citri* pv. *mangiferaeindicae*, one strain of *X. citri* pv. *anacardii*, three strains of *X. oryzae* pv. *oryzicola*, and used 51 other publically available *Xanthomonas* genomes for a total of 72 whole

genome sequences (Table 1). *Stenotrophomonas maltophilia* strain R551–3 and *Xylella fastidiosa* strains 9a5c and Temecula1 were used as outgroups for further analyses [19–21]. Annotation revealed from 3209 to 5405 coding sequences (CDS) per *Xanthomonas* genome (Additional file 1). Among *Xanthomonas* strains responsible for CBB, chromosome size ranged from 4,957,446 bp in strain CFBP1815 to 5,517,999 bp in strain CFBP6992, with an average GC content ranging from 64.3 to 64.9%. The phylogeny of strains was assessed based on the amino acid sequences of all annotated CDS using CVTree (Fig. 1). The overall topology of this tree was congruent with previous *Xanthomonas* phylogenies [11, 12, 22]. As described previously, strains responsible for CBB are distributed into four distinct genetic lineages belonging to two different species, *X. citri* and *X. phaseoli* [12, 13, 16, 23].

### Genome expansion occurred during the evolution of *Xanthomonas*

To identify the genes shared by different clades of *Xanthomonas*, we constructed an orthology matrix using OrthoMCL (Additional file 2). Based on this orthology matrix, we performed a parsimony approach to infer gene gains and losses at each branch of the phylogenetic tree (Fig. 1). We did not take into account events occurring on the most distal branches to reduce the bias due to the difference of quality between sequenced genomes. At every branch, one to several hundreds of genes were either gained or lost. A general observation was that gene gains were higher than gene losses, suggesting that genome expansion occurred during the evolution of *Xanthomonas* (Fig. 1). Only four cases of genome reduction were observed (i) at the origin of the *Xylella* genus, (ii and iii) along two consecutive branches before and after the split between the *X. oryzae* species and the *X. vasculorum* and *X. musacearum* species, and (iv) at the origin of *X. citri* pv. *malvacearum*. The largest gene gain (418) was observed at the origin of *Xanthomonas* phylogenetic group 2 as defined by Young et al. [24], while the largest gene loss (819) was observed at the origin of the *Xylella* genus. Few gene losses (9 to 32) were observed before the diversification of each of the four genetic lineages involved in CBB. Of those, the NF1 lineage was the one which gained the most genes (271) followed by the NF2 (225), NF3 (108) and fuscans (83) lineages, respectively.

### The pan and core genomes of *Xanthomonas* reveal extensive horizontal gene transfers between strains pathogenic on common bean

Individuals that are closely related to each other typically share more orthologs than unrelated individuals. Therefore, groups of closely related individuals tend to have a smaller pan genome and a larger core genome than

**Table 1** Informations on the sequenced strains used in this study

| Identifier  | Species/Pathovar                             | Strain     | Host of isolation                               | Country and date of isolation | Contigs | Total size (bp) | GC%   | Accession number | Reference                   |
|-------------|----------------------------------------------|------------|-------------------------------------------------|-------------------------------|---------|-----------------|-------|------------------|-----------------------------|
| stenma5513  | <i>Stenotrophomonas maltophilia</i>          | R551-3     | <i>Populus trichocarpa</i>                      | na                            | 1       | 4,573,969       | 66.29 | GCA_000020665.1  | Taghavi et al., 2008 [21]   |
| albilliPC73 | <i>X. albilineans</i>                        | GPE PC73   | <i>Saccharum</i> spp. cv. H63-1418.             | Guadeloupe (France), 2003     | 1       | 3,768,695       | 62.97 | GCA_000087965.1  | Pieretti et al., 2009 [77]  |
| aberra6865  | <i>X. campestris</i> pv. <i>campestris</i>   | CFBP6865   | <i>Brassica oleracea</i> var. <i>capitata</i> . | Australia, 1975               | 4       | 5,169,518       | 64.93 | Unpublished      | Noël, comm. pers            |
| barbar5825  | <i>X. campestris</i> pv. <i>barbareae</i>    | CFBP5825R  | <i>Barbarea vulgaris</i>                        | USA, 1939                     | 2       | 5,055,186       | 65.09 | ATNQ000000000    | Roux et al., 2015 [78]      |
| campes8004  | <i>X. campestris</i> pv. <i>campestris</i>   | 8004       | <i>Brassica oleracea</i> var. <i>botrytis</i> . | United Kingdom, 1958          | 1       | 5,148,708       | 64.95 | GCA_000012105.1  | Qian et al., 2005 [79]      |
| campesAT33  | <i>X. campestris</i> pv. <i>campestris</i>   | ATCC 33913 | <i>Brassica oleracea</i> var. <i>gemmifera</i>  | United Kingdom, 1957          | 1       | 5,076,188       | 65.06 | GCA_000007145.1  | da Silva et al., 2002 [80]  |
| campesB100  | <i>X. campestris</i> pv. <i>campestris</i>   | B100       | na                                              | na                            | 1       | 5,079,002       | 65.04 | GCA_000070605.1  | Vorhöfner et al., 2008 [81] |
| incana2527  | <i>X. campestris</i> pv. <i>incanae</i>      | CFBP2527R  | <i>Matthiola incana</i>                         | USA, 1950                     | 1       | 4,926,205       | 65.12 | ATNO000000000    | Roux et al., 2015 [78]      |
| incana1606  | <i>X. campestris</i> pv. <i>incanae</i>      | CFBP1606R  | <i>Matthiola incana</i>                         | France, 1974                  | 1       | 4,967,651       | 65.17 | ATNN000000000    | Roux et al., 2015 [78]      |
| musace4381  | <i>X. campestris</i> pv. <i>musacearum</i>   | NCPB4381   | <i>Musa</i> sp.                                 | Uganda, 2005                  | 115     | 4,810,038       | 61.96 | ACHT000000000    | Studholme et al., 2010 [82] |
| raphan5828  | <i>X. campestris</i> pv. <i>raphani</i>      | CFBP5828R  | <i>Raphanus sativus</i>                         | USA, na                       | 2       | 4,912,709       | 65.35 | ATNP000000000    | Roux et al., 2015 [78]      |
| raphan756C  | <i>X. campestris</i> pv. <i>raphani</i>      | 756C       | <i>Brassica oleracea</i> var. <i>capitata</i>   | east asia, na                 | 1       | 4,941,214       | 65.28 | GCA_000221965.1  | Bogdanove et al., 2011 [83] |
| cassav4642  | <i>X. cassavae</i>                           | CFBP4642   | <i>Manihot esculenta</i>                        | Malawi, 1951                  | 7       | 5,278,192       | 65.22 | GCA_000454545.1  | Bolot et al., 2013 [84]     |
| axanac2913  | <i>X. citri</i> pv. <i>anacardii</i>         | CFBP2913   | <i>Mangifera indica</i>                         | Brazil, na                    | 1       | 5,203,496       | 64.57 | CP024057         | this study                  |
| axaura1035  | <i>X. citri</i> pv. <i>aurantifolii</i>      | ICPB10535  | <i>Citrus aurantiifolia</i>                     | Brazil, na                    | 351     | 5,060,896       | 64.81 | ACPY000000000    | Moreira et al., 2010 [85]   |
| axcitr1083  | <i>X. citri</i> pv. <i>citri</i>             | FDC1083    | <i>Citrus reticulata</i>                        | Brazil, 1980                  | 1       | 5,218,314       | 64.72 | CCVZ010000000    | Gordon et al., 2015 [86]    |
| axcitr1902  | <i>X. citri</i> pv. <i>citri</i>             | JF90-2     | <i>Citrus aurantiifolia</i>                     | Oman, 1986                    | 1       | 5,251,225       | 64.66 | CCWA010000000    | Gordon et al., 2015 [86]    |
| axcitr1238  | <i>X. citri</i> pv. <i>citri</i>             | JJ238-24   | <i>Citrus aurantiifolia</i>                     | Thailand, 1989                | 1       | 5,282,622       | 64.78 | CCVX010000000    | Gordon et al., 2015 [86]    |
| axcitrLC80  | <i>X. citri</i> pv. <i>citri</i>             | LC80       | <i>Citrus reticulata</i> x <i>C. sinensis</i>   | Mali, 2006                    | 1       | 5,229,127       | 64.7  | CCWJ010000000    | Gordon et al., 2015 [86]    |
| axcitrLE20  | <i>X. citri</i> pv. <i>citri</i>             | LE20-1     | <i>Citrus aurantiifolia</i>                     | Ethiopia, 2008                | 1       | 5,309,240       | 64.71 | CCWK010000000    | Gordon et al., 2015 [86]    |
| axcitr9322  | <i>X. citri</i> pv. <i>citri</i>             | LMG9322    | <i>Citrus aurantiifolia</i>                     | Florida, 1986                 | 1       | 5,194,131       | 64.75 | CCVY010000000    | Gordon et al., 2015 [86]    |
| axcitr306   | <i>X. citri</i> pv. <i>citri</i>             | 306        | na                                              | na                            | 1       | 5,175,554       | 64.77 | GCA_000007165.1  | da Silva et al., 2002 [80]  |
| phafus1815  | <i>X. citri</i> pv. <i>fuscans</i> (fuscans) | CFBP1815   | <i>Phaseolus</i> sp                             | Greece, 1978                  | 1       | 4,957,446       | 64.82 | GCA_900234415    | this study                  |
| phafus4834  | <i>X. citri</i> pv. <i>fuscans</i> (fuscans) | 4834-R     | <i>Phaseolus vulgaris</i> cv. Michelet          | France, 1998                  | 1       | 4,981,995       | 64.81 | FO681494         | Darrasse et al., 2013 [26]  |
| phafus6166  | <i>X. citri</i> pv. <i>fuscans</i> (fuscans) | CFBP6166   | <i>Phaseolus vulgaris</i>                       | South Africa, 1963            | 1       | 4,987,587       | 64.8  | GCA_900234455    | this study                  |
| phafus6960  | <i>X. citri</i> pv. <i>fuscans</i> (fuscans) | CFBP6960   | <i>Phaseolus vulgaris</i>                       | Reunion island, France, 2000  | 1       | 4,992,131       | 64.79 | GCA_900234515    | this study                  |
| phafus6970  | <i>X. citri</i> pv. <i>fuscans</i> (fuscans) | CFBP6970   | <i>Phaseolus</i> sp                             | USA, 1990                     | 1       | 5,006,656       | 64.9  | GCA_900234505    | this study                  |
| pha2GL6988  | <i>X. citri</i> pv. <i>fuscans</i> (NF2)     | CFBP6988   | <i>Phaseolus vulgaris</i> cv. marla             | Reunion island, France, 2000  | 1       | 5,132,433       | 64.64 | Deposited        | this study                  |

**Table 1** Informations on the sequenced strains used in this study (Continued)

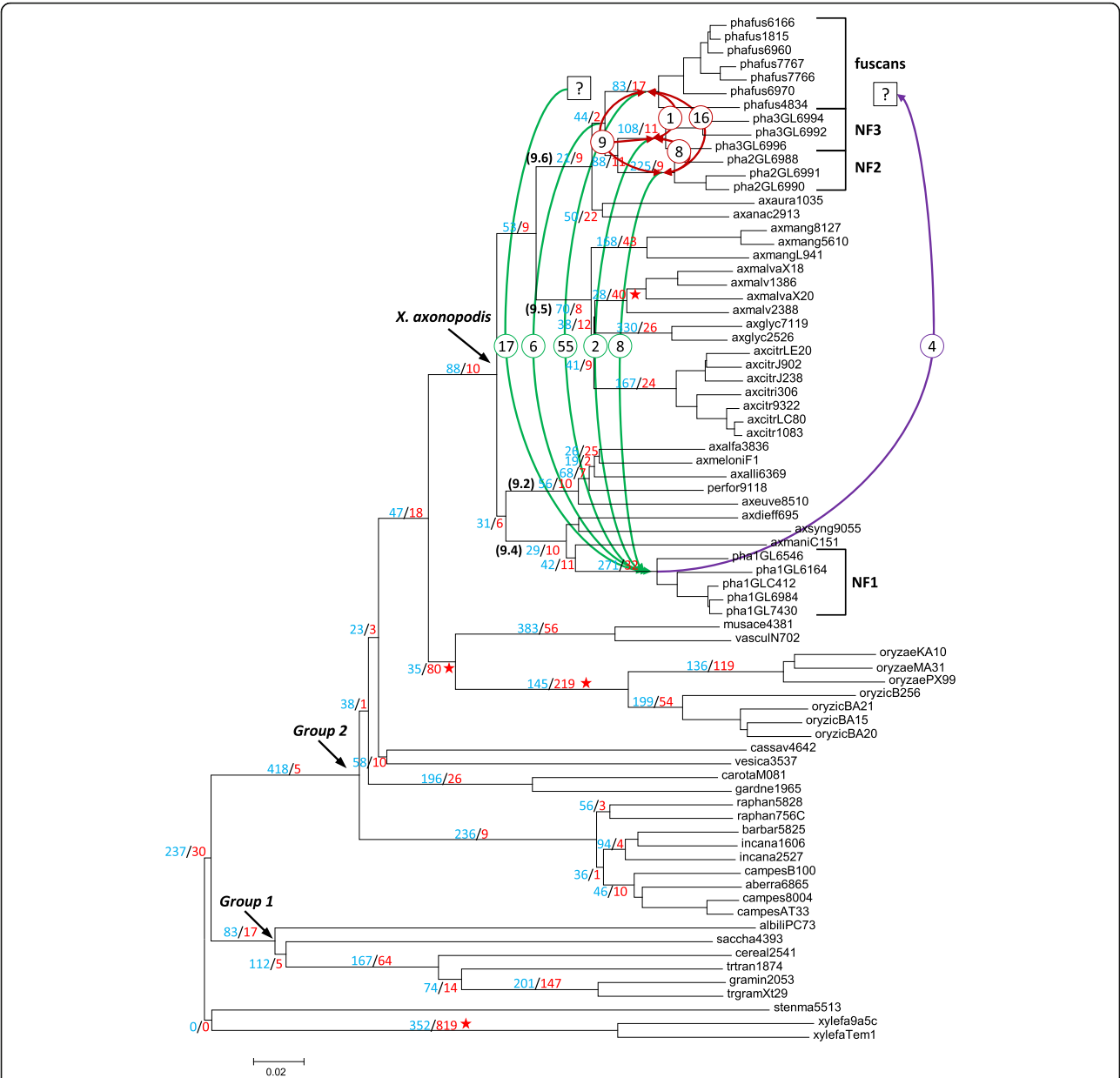
| Identifier | Species/Pathovar                                 | Strain      | Host of isolation                   | Country and date of isolation | Contigs | Total size (bp) | GC%   | Accession number | Reference                    |
|------------|--------------------------------------------------|-------------|-------------------------------------|-------------------------------|---------|-----------------|-------|------------------|------------------------------|
| pha2GL6990 | <i>X. citri</i> pv. <i>fuscans</i> (NF2)         | CFBP6990    | <i>Phaseolus vulgaris</i> cv. marla | Reunion island, France, 2000  | 1       | 5,124,653       | 64.63 | GCA_900234485    | this study                   |
| pha2GL6991 | <i>X. citri</i> pv. <i>fuscans</i> (NF2)         | CFBP6991    | <i>Phaseolus vulgaris</i> cv. marla | Reunion island, France, 2000  | 1       | 5,342,565       | 64.35 | GCA_900234525    | this study                   |
| pha3GL6992 | <i>X. citri</i> pv. <i>fuscans</i> (NF3)         | CFBP6992    | <i>Phaseolus vulgaris</i> cv. marla | Reunion island, France, 2000  | 1       | 5,517,999       | 64.48 | GCA_900234495    | this study                   |
| pha3GL6994 | <i>X. citri</i> pv. <i>fuscans</i> (NF3)         | CFBP6994    | <i>Phaseolus vulgaris</i>           | Tanzania, 1990                | 1       | 5,250,266       | 64.57 | GCA_900234565    | this study                   |
| pha3GL6996 | <i>X. citri</i> pv. <i>fuscans</i> (NF3)         | CFBP6996    | <i>Phaseolus vulgaris</i> cv. marla | Reunion island, France, 2000  | 4       | 5,095,420       | 64.72 | AVET00000000     | this study                   |
| phafus7766 | <i>X. citri</i> pv. <i>fuscans</i> (fuscans)     | CFBP7766    | <i>Phaseolus vulgaris</i>           | Cameroon, 2009                | 1       | 5,185,891       | 64.63 | GCA_900234475    | this study                   |
| phafus7767 | <i>X. citri</i> pv. <i>fuscans</i> (fuscans)     | CFBP7767    | <i>Phaseolus vulgaris</i>           | Cameroon, 2009                | 1       | 5,107,388       | 64.7  | GCA_900234465    | this study                   |
| axglyc2526 | <i>X. citri</i> pv. <i>glycines</i>              | CFBP2526    | <i>Glycine hispida</i>              | Sudan, 1956                   | 4       | 5,255,152       | 64.63 | GCA_000495275.1  | Darrasse et al., 2013 [26]   |
| axglyc7119 | <i>X. citri</i> pv. <i>glycines</i>              | CFBP7119    | <i>Glycine max</i>                  | Brazil, 1981                  | 4       | 5,521,783       | 64.39 | GCA_000488895.1  | Darrasse et al., 2013 [26]   |
| axmalv1386 | <i>X. citri</i> pv. <i>malvacearum</i>           | GSPB1386    | na                                  | na                            | 127     | 4,991,411       | 64.7  | GCA_000309905.1  | Hainan University            |
| axmalv2388 | <i>X. citri</i> pv. <i>malvacearum</i>           | GSPB2388    | na                                  | na                            | 61      | 5,127,016       | 64.53 | GCA_000309925.1  | Hainan University            |
| axmalvaX18 | <i>X. citri</i> pv. <i>malvacearum</i>           | X18         | <i>Gossypium</i> spp.               | Burkina Faso, 1980s           | 4       | 4,993,692       | 64.7  | ATMA000000000    | Cunnac et al., 2013 [87]     |
| axmalvaX20 | <i>X. citri</i> pv. <i>malvacearum</i>           | X20         | <i>Gossypium</i> spp.               | Burkina Faso, 1980s           | 4       | 5,219,648       | 64.49 | ATMB000000000    | Cunnac et al., 2013 [87]     |
| axmang5610 | <i>X. citri</i> pv. <i>mangiferaeindicae</i>     | LG56-10     | <i>Mangifera indica</i>             | na                            | 4       | 5,281,952       | 64.56 | PEBY000000000    | this study                   |
| axmang8127 | <i>X. citri</i> pv. <i>mangiferaeindicae</i>     | LG81-27     | <i>Mangifera indica</i>             | na                            | 6       | 5,209,594       | 64.68 | PEBZ000000000    | this study                   |
| axmangL941 | <i>X. citri</i> pv. <i>mangiferaeindicae</i>     | LMG941      | <i>Mangifera indica</i>             | India, 1957                   | 195     | 5,144,323       | 64.84 | CAHO01000001     | Midha et al., 2012 [88]      |
| axalfa3836 | <i>X. euvesicatoria</i> pv. <i>alfalfae</i>      | CFBP3836    | <i>Medicago sativa</i>              | Sudan, na                     | 6       | 5,081,438       | 64.74 | GCA_000488955.1  | Jacques et al., 2013 [3, 22] |
| axall6369  | <i>X. euvesicatoria</i> pv. <i>alii</i>          | CFBP6369    | <i>Allium cepa</i>                  | Reunion island, France, 1996  | 3       | 5,427,488       | 64.36 | GCA_000730305.1  | Gagnevin et al., 2014 [89]   |
| axmelonF1  | <i>X. euvesicatoria</i> pv. <i>citrumelonis</i>  | F1          | <i>Citrus</i> sp.                   | FL, USA, 1984                 | 1       | 4,967,469       | 64.91 | GCA_000225915.1  | Jalan et al., 2011 [90]      |
| axeuve8510 | <i>X. euvesicatoria</i> pv. <i>euvesicatoria</i> | 85-10       | <i>Capsicum annuum</i>              | FL, USA, 1985                 | 1       | 5,178,466       | 64.74 | GCA_000009165.1  | Thieme et al., 2005 [91]     |
| perfor9118 | <i>X. euvesicatoria</i> pv. <i>perforans</i>     | 91-118      | <i>Lycopersicon esculentum</i>      | FL, USA, na                   | 291     | 5,296,241       | 65.04 | AEQW000000000    | Potnis et al., 2011 [92]     |
| gardne1965 | <i>X. gardneri</i>                               | ATCC19865   | <i>Lycopersicon esculentum</i>      | Yugoslavia, 1953              | 552     | 5,594,687       | 63.68 | AEQX000000000    | Potnis et al., 2011 [92]     |
| carotaM081 | <i>X. hortorum</i> pv. <i>carotae</i>            | M081        | <i>Daucus carota</i>                | OR, USA, na                   | 154     | 5,039,269       | 63.83 | GCA_000505565.1  | Kimbril et al., 2011 [93]    |
| onyzaeKA10 | <i>X. onyzae</i> pv. <i>onyzae</i>               | KACC10331   | <i>Onyza sativa</i>                 | Korea, na                     | 1       | 4,941,439       | 63.69 | GCA_000007385.1  | Lee et al., 2005 [94]        |
| onyzaeMA31 | <i>X. onyzae</i> pv. <i>onyzae</i>               | MAFF 311018 | <i>Onyza sativa</i>                 | Japan, na                     | 1       | 4,940,217       | 63.7  | GCA_000010025.1  | unpublished                  |
| onyzaePX99 | <i>X. onyzae</i> pv. <i>onyzae</i>               | PXO99A      | <i>Onyza sativa</i>                 | Philippines, na               | 1       | 5,240,075       | 63.63 | GCA_000019585.1  | Salzberg et al., 2008 [95]   |
| onyzacBA15 | <i>X. onyzae</i> pv. <i>onyzicola</i>            | BA115       | <i>Onyza sativa</i>                 | Burkina Faso, 2009            | 1       | 4,315,327       | 64.16 | GCA_0002189395.1 | this study                   |

**Table 1** Informations on the sequenced strains used in this study (Continued)

| Identifier | Species/Pathovar                                   | Strain    | Host of isolation                   | Country and date of isolation | Contigs | Total size (bp) | GC%   | Accession number | Reference                    |
|------------|----------------------------------------------------|-----------|-------------------------------------|-------------------------------|---------|-----------------|-------|------------------|------------------------------|
| oryzicBA20 | <i>X. oryzae</i> pv. <i>oryzicola</i>              | BAI20     | <i>Oryza sativa</i>                 | Burkina Faso, 2009            | 1       | 4,498,643       | 64.05 | GCA_002189435.1  | this study                   |
| oryzicBA21 | <i>X. oryzae</i> pv. <i>oryzicola</i>              | BAI21     | <i>Oryza sativa</i>                 | Burkina Faso, 2009            | 1       | 4,419,870       | 64.07 | GCA_002189465.1  | this study                   |
| oryzicB256 | <i>X. oryzae</i> pv. <i>oryzicola</i>              | BLS256    | <i>Oryza sativa</i>                 | Philippines, 1984             | 1       | 4,831,739       | 64.05 | AAQN01000001     | Bogdanove et al., 2011 [83]  |
| axdieff695 | <i>X. phaseoli</i> pv. <i>dieffenbachiae</i>       | LMG695    | <i>Anthurium andreanum</i>          | Brazil, 1965                  | 1       | 5,037,357       | 64.88 | GCA_001564415.1  | Robène et al., 2016 [96]     |
| axmaniC151 | <i>X. phaseoli</i> pv. <i>manihotis</i>            | CIO151    | <i>Manihot</i> sp.                  | Colombia, na                  | 36      | 5,154,532       | 64.7  | GCA_000265845.1  | Rodriguez et al., 2012 [97]  |
| pha1GLC412 | <i>X. phaseoli</i> pv. <i>phaseoli</i> (NF1)       | CFBP412   | <i>Phaseolus vulgaris</i>           | USA, na                       | 1       | 5,028,351       | 64.94 | GCA_900234435    | this study                   |
| pha1GL6164 | <i>X. phaseoli</i> pv. <i>phaseoli</i> (NF1)       | CFBP6164  | <i>Phaseolus vulgaris</i>           | Romania, 1996                 | 1       | 5,319,061       | 64.64 | GCA_900234535    | this study                   |
| pha1GL6546 | <i>X. phaseoli</i> pv. <i>phaseoli</i> (NF1)       | CFBP6546  | <i>Phaseolus vulgaris</i>           | USA, na                       | 4       | 4,994,665       | 64.92 | Deposited        | this study                   |
| pha1GL6984 | <i>X. phaseoli</i> pv. <i>phaseoli</i> (NF1)       | CFBP6984  | <i>Phaseolus vulgaris</i>           | Reunion island, France, 2000  | 1       | 5,106,663       | 64.82 | GCA_900234425    | this study                   |
| pha1GL7430 | <i>X. phaseoli</i> pv. <i>phaseoli</i> (NF1)       | CFBP7430  | <i>Phaseolus vulgaris</i> cv. CFO16 | Iran, 2006                    | 1       | 5,085,978       | 64.89 | GCA_900234445    | this study                   |
| axsyng9055 | <i>X. phaseoli</i> pv. <i>syngonii</i> (NF1)       | LMG9055   | <i>Syngonium podophyllum</i>        | na                            | 6       | 5,010,905       | 64.82 | GCA_001640215.1  | Robène et al., 2016 [96]     |
| saccha4393 | <i>X. sacchari</i>                                 | NCPB4393  | <i>Musa</i> sp.                     | Tanzania, 2007                | 470     | 4,955,099       | 69.03 | AGDB000000000    | Studholme et al., 2011 [98]  |
| cereal2541 | <i>X. translucens</i> pv. <i>cerealis</i>          | CFBP2541  | <i>Bromus inermis</i>               | USA, 1941                     | 10      | 4,524,870       | 67.37 | GCA_000807145.1  | Pesce et al., 2015a [99]     |
| trgramXt29 | <i>X. translucens</i> pv. <i>graminis</i>          | ART-Xtg29 | <i>Lolium multiflorum</i>           | Switzerland, na               | 788     | 4,203,855       | 68.58 | ANGG000000000    | Wichman et al., 2013 [101]   |
| gramin2053 | <i>X. translucens</i> pv. <i>graminis</i>          | CFPB2053  | <i>Dactylis glomerata</i>           | Switzerland, 1973             | 2       | 4,344,936       | 68.33 | LHSI000000000    | Pesce et al., 2015b [100]    |
| trtran1874 | <i>X. translucens</i> pv. <i>translucens</i>       | DSM18974  | <i>Hordeum vulgare</i>              | USA, 1958                     | 551     | 4,550,921       | 67.67 | GCA_000331775.1  | unpublished                  |
| vasculN702 | <i>X. vasicola</i> pv. <i>vasculorum</i>           | NCPB702   | <i>Saccharum officinarum</i>        | Zimbabwe, 1959                | 97      | 5,491,457       | 56.86 | ACHS000000000    | Studholme et al., 2010 [82]  |
| vesica3537 | <i>X. vesicatoria</i>                              | ATCC35937 | <i>Lycopersicon esculentum</i>      | New Zealand, 1955             | 296     | 5,567,561       | 64.06 | AEQV010000000    | Potnis et al., 2011 [92]     |
| xylefaTem1 | <i>Xylella fastidiosa</i> subsp. <i>fastidiosa</i> | Temecula1 | <i>Vitis vinifera</i>               | USA, na                       | 1       | 2,519,802       | 51.77 | GCA_000007245.1  | van Sluys et al., 2003 [102] |
| xylefa9a5c | <i>Xylella fastidiosa</i> subsp. <i>pauca</i>      | 9a5c      | <i>Citrus sinensis</i> cv. Valencia | Brazil, 1992                  | 1       | 2,679,306       | 52.67 | GCA_000006725.1  | Simpson et al., 2000 [19]    |

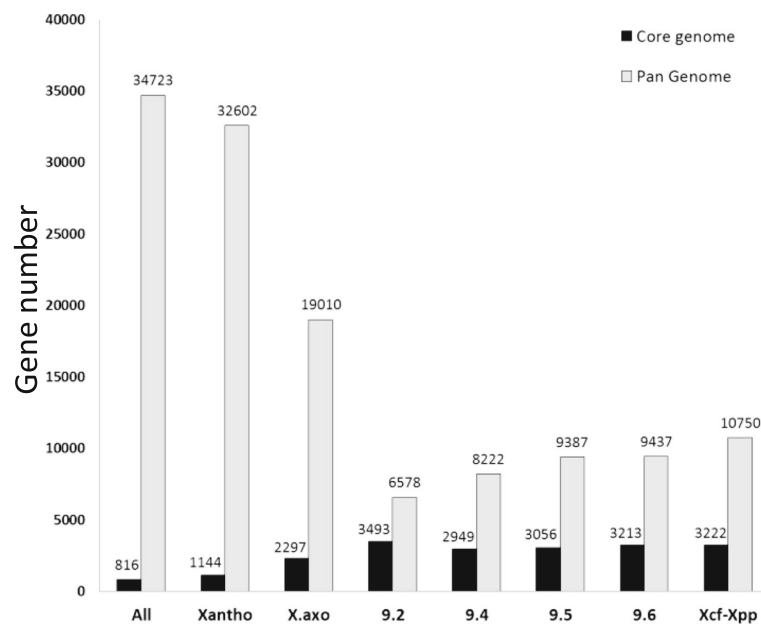
na not available





**Fig. 1** Phylogeny of *Xanthomonas* strains used in this study. Phylogeny of *Xanthomonas* strains used in this study with indication of gene gains and losses. The phylogenetic tree is based on whole genome analysis using CVTree [66] with default parameters. Strain aliases are described in Table 1. *Stenotrophomonas* and *Xylella* genomes have been used as outgroups. *Xanthomonas* main phylogenetic groups 1 and 2 [24] and the *X. axonopodis* species complex [9, 10] are indicated by arrows. Groups 9.2, 9.4, 9.5 and 9.6 [11] are indicated in brackets. Fuscans, NF2, NF3 and NF1 refer to the four genetic lineages of strains responsible for CBB. A parsimony approach was performed to infer gene gains (blue) and losses (red) at levels higher than the pathovar rank, and numbers are displayed at each branch. Red stars highlight cases where gene loss was greater than gene gain. Curved arrows represent horizontal gene transfers (HGT) retrieved by Ks analysis on alignments of 115 candidate genes for bean specificity, with HGT from *X. citri* pv. *fuscans* to *X. phaseoli* pv. *phaseoli* in green, HGT from *X. phaseoli* pv. *phaseoli* to *X. citri* pv. *fuscans* in purple, and HGT between *X. citri* pv. *fuscans* lineages in red. Numbers in circles correspond to the numbers of candidate genes involved for each HGT. Question marks indicate events for which the origin or end of the HGT was not precise enough to assign any particular lineage

groups of more divergent individuals. As such, the pan and core genomes for the 72 *Xanthomonas* strains comprised 32,602 and 1144 CDS, respectively, while the pan and core genomes for the 75 strains including the outgroups comprised 34,723 and 816 CDS, respectively (Fig. 2). Similarly, each Rademaker group alone, i.e. 9.2, 9.4, 9.5 and 9.6, had a smaller pan genome (6578, 8222, 9387 and 9437 CDS, respectively) and a larger core genome (3493, 2949, 3056 and 3213 CDS, respectively) than the *X.*



**Fig. 2** The core- and pan-genome of *Xanthomonas*. Gene numbers correspond to the number of ortholog groups retrieved for each group of strains. All: all strains used in this study ( $n = 75$ ); Xantho: strains from the *Xanthomonas* genus ( $n = 72$ ); X.axo: strains from the *Xanthomonas axonopodis* species complex ( $n = 44$ ). 9.2, 9.4, 9.5, 9.6: strains belonging to rep-PCR groups 9.2 ( $n = 5$ ), 9.4 ( $n = 8$ ), 9.5 ( $n = 16$ ) and 9.6 ( $n = 15$ ), respectively, as defined in Rademaker et al. [103]. Xcf-Xpp: strains pathogenic on common bean belonging to *Xanthomonas citri* pv. *fuscans* or *Xanthomonas phaseoli* pv. *phaseoli* ( $n = 18$ )

*axonopodis* species complex, which had pan and core genomes of 19,010 and 2297 CDS, respectively. Strikingly, when grouping strains responsible for CBB belonging to *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli*, both the pan and core genomes (10,750 and 3222 CDS, respectively) were larger than the pan and core genomes from groups 9.4 (8222 and 2949 CDS, respectively) or 9.6 (9437 and 3213 CDS, respectively) (Fig. 2). Thus, strains responsible for CBB, although being phylogenetically diverse, had more genes in common than they had with other strains belonging to their respective clades, which was suggestive of extensive HGT among these strains. This result was reminiscent of previous comparative analyses showing that dozens of genes have been horizontally transferred between the fuscans and NF1 lineages [17].

#### Strains pathogenic on common bean share 115 CDS presenting unique characteristics

To search for genes potentially involved in the convergence between *X. phaseoli* pv. *phaseoli* (i.e. the NF1 lineage) and *X. citri* pv. *fuscans* (i.e. the NF2, NF3 and fuscans lineages) to infect common bean, we performed a combination of four different analyses. First, within the OrthoMCL matrix, we searched for CDS specifically present in the genomes of CBB agents and absent from any other *Xanthomonas* genome, or present in the genomes from all *Xanthomonas* but not in the genomes of

CBB agents. No CDS was retrieved by this analysis. We also searched for CDS specifically present or absent when grouping the NF1 lineage to each of the NF2, NF3, or fuscans lineages. Only one CDS was specifically retrieved in the NF1 and fuscans lineages.

Second, we used the results from the CDS gains and losses approach described above to search for genes shared by all strains from *X. phaseoli* pv. *phaseoli* and *X. citri* pv. *fuscans*, and gained in the ancestor of one pathovar or the other. This approach unveiled nine CDS shared by all strains responsible for CBB, and gained in either *X. phaseoli* pv. *phaseoli* or *X. citri* pv. *fuscans* (Table 2). We also searched for CDS shared by the NF1 and each of the NF2, NF3, or fuscans lineage and gained in at least one of these lineages. Four CDS were shared by the NF1 and NF2 lineages, or the NF1 and NF3 lineages, while three CDS were shared by the NF1 and fuscans lineages.

Third, we used a phylogenetic approach to search for genes for which strains from *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* formed a monophyletic group. For this, we constructed phylogenetic trees on 3202 CDS present in every *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* strain and in at least one additional strain from group 9.2 or 9.4 and one other strain from group 9.5 or 9.6. The additional strains from groups 9.2, 9.4, 9.5 and 9.6 were located inbetween *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* (Fig. 1). Thus, CDS found as



**Table 2** Numbers of CDS presenting similarities among the lineages of CBB agents

| Lineages studied    | Presence/absence <sup>a</sup> | Gained <sup>b</sup> | Monophyletic <sup>c</sup> | 24-mers <sup>d</sup> |
|---------------------|-------------------------------|---------------------|---------------------------|----------------------|
| NF1/NF2/NF3/fuscans | 0                             | 9                   | 28                        | 108                  |
| NF1/NF2             | 0                             | 4                   | 9                         | 33                   |
| NF1/NF3             | 0                             | 4                   | 5                         | 28                   |
| NF1/fuscans         | 1                             | 3                   | 105                       | 231                  |
| Total               | 1                             | 20                  | 147                       | 400                  |

<sup>a</sup>CDS specifically present or absent in all the lineages studied compared to all other *X. axonopodis* strains

<sup>b</sup>CDS present in all the lineages studied and gained in a least one of these lineages

<sup>c</sup>CDS monophyletic for the lineages studied

<sup>d</sup>CDS containing 24-mers specifically present or absent in all the lineages studied compared to all other *X. axonopodis* strains

monophyletic for CBB strains could be potential traces of HGT between both pathogens. This approach unveiled 28 CDS for which the four genetic lineages formed a monophyletic group, suggesting that they were horizontally transferred among these lineages (Table 2). Nine CDS were specifically monophyletic for the NF1 and NF2 lineages, five for the NF1 and NF3 lineages, and 105 for the NF1 and fuscans lineages, suggesting that most horizontal transfers occurred among the NF1 and fuscans lineages.

Finally, we used the SkIf tool [25] on the 72 *Xanthomonas* genomes to search for genes containing short 24-bp sequences (24-mers) specific to strains responsible for CBB, or alternatively genes from strains responsible for CBB lacking 24-mers present in all other strains from the *X. axonopodis* species complex. In all, we identified 108 CDS containing 24-mers either specifically present or absent from the four lineages (Table 2). Moreover, 33 CDS contained 24-mers specific for the NF1 and NF2 lineages, 28 for the NF1 and NF3 lineages and 231 for the NF1 and fuscans lineages. Similarly to the analysis based on phylogeny, this analysis based on k-mers pointed an overrepresentation of CDS with specific 24-mers shared by the NF1 and fuscans lineages compared to NF2 and NF3 lineages.

Together, these four analyses unveiled respectively 0, 9, 28, or 109 CDS presenting features unique to CBB agents. The analysis based on presence/absence seemed to be too stringent for unveiling any CDS, while the analysis based on k-mers was the most sensitive, suggesting that SkIf was an appropriate tool for finding common traits shared by phylogenetically distant strains. Most of these CDS were found redundantly by two or more analyses, for a total of 115 non-redundant CDS (Table 3). The most represented functions encoded by these 115 predicted CDS were hypothetical proteins (26 CDS), followed by membrane-related proteins (10 CDS), two-component system proteins (six CDS), putative secreted proteins (five CDS), reductases (five CDS), RNA-related proteins (five CDS), Type III secretion system-related proteins (five CDS), TonB-dependent proteins (four CDS), Type IV secretion system-related

proteins (three CDS), Type VI secretion system-related proteins (three CDS), DNA-related proteins (three CDS) and transcription regulators (three CDS) (Table 3).

We hypothesized that the CDS potentially involved in the specific adaptation to common bean should bear nonsynonymous polymorphisms specific to *Xanthomonas* strains pathogenic on common bean. Analysis of the alignments for the 115 candidate CDS highlighted 44 CDS with nonsynonymous sites retrieved exclusively in *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* (Table 3). More than one third of these CDS (16/44) encoded hypothetical proteins. Among the other CDS, three encoded type IV secretion system proteins TrbI, TraG and TraF, two encoded putative secreted proteins, two encoded type III secretion system proteins XopA and XopAD, two encoded DNA topoisomerases, and others encoded proteins of various functions (Table 3).

#### Specificity to common bean is associated with successive waves of horizontal gene transfers

Strain CFBP6546 from the NF1 lineage was used as reference for further analyses. Its genome contained one chromosome and three extrachromosomal plasmids formerly described as plasmid a, plasmid b and plasmid c in strain 4834-R [26]. Most candidate genes (95/115) were located on the chromosome, while 20 were located on the plasmid a (Fig. 3). This corresponds to a density of one candidate gene per 50.9 kbp in the chromosome, and one per 3.5 kbp for the plasmid a, while none were retrieved in plasmids b or c. Interestingly, all the CDS found in plasmid a contained specific nonsynonymous sites (Table 3). Thus, plasmid a appeared as an important vector of genes involved in the adaptation to common bean. Another observation was that the chromosome contained regions with various 24-mers shared by the NF1 lineage and any of the fuscans, NF2 or NF3 lineages (in green, blue or black in Fig. 3, respectively). This suggests that the regions shared by the NF1 and the other lineages diverged since the split between the NF2, NF3 and fuscans lineages. By contrast to what was observed for the chromosome, all specific 24-mers found in plasmid a were simultaneously shared

**Table 3** Overview of the 115 genes putatively involved in bean specificity

| Identifier <sup>a</sup> | Accession number <sup>b</sup> | Predicted function                      | Gene          | Gained | Monophyletic | 24-mers | Recombinant | HGT | Atypical GC% | Nonsynonymous sites | Aritua et al. [17] <sup>c</sup> |
|-------------------------|-------------------------------|-----------------------------------------|---------------|--------|--------------|---------|-------------|-----|--------------|---------------------|---------------------------------|
| 1                       | m00100560                     | TonB-dependent transporter              | -             | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 2                       | m00100580                     | type III effector                       | <i>avrBs2</i> | -      | -            | Yes     | Yes         | Yes | -            | -                   | -                               |
| 3                       | m00100590                     | hypothetical protein                    | -             | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 4                       | m00101230                     | two component system sensor protein     | -             | -      | -            | Yes     | -           | -   | -            | -                   | -                               |
| 5                       | m00101980                     | two component system protein            | <i>glhG</i>   | -      | -            | Yes     | -           | -   | -            | -                   | -                               |
| 6                       | m00102200                     | two component system response regulator | -             | -      | Yes          | Yes     | -           | Yes | -            | -                   | Yes                             |
| 7                       | m00104250                     | type III secretion system protein       | <i>hrpF</i>   | -      | -            | Yes     | Yes         | Yes | -            | -                   | -                               |
| 8                       | m00104520                     | type III effector                       | <i>xopA</i>   | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 9                       | m00104530                     | lytic transglycosylase-like protein     | <i>hpaH</i>   | -      | -            | Yes     | -           | Yes | -            | Yes                 | Yes                             |
| 10                      | m00104540                     | hypothetical protein                    | -             | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 11                      | m00104640                     | diguanylate cyclase                     | -             | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 12                      | m00104690                     | 4-alpha-glucanotransferase              | <i>malQ</i>   | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 13                      | m00105290                     | threonine aldolase                      | -             | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 14                      | m00105330                     | flavin reductase                        | -             | -      | Yes          | Yes     | -           | Yes | -            | Yes                 | -                               |
| 15                      | m00105340                     | anthranilate phosphoribosyl transferase | <i>trpD</i>   | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 16                      | m00105390                     | S-adenosylmethionine decarboxylase      | <i>speD</i>   | -      | -            | Yes     | -           | Yes | Yes          | -                   | -                               |
| 17                      | m00105750                     | hypothetical membrane protein           | -             | -      | -            | Yes     | -           | -   | -            | -                   | -                               |
| 18                      | m00107500                     | TonB-dependent transporter              | -             | -      | -            | Yes     | Yes         | Yes | -            | -                   | -                               |
| 19                      | m00108560                     | general secretion pathway protein D     | <i>xpsD</i>   | -      | Yes          | Yes     | Yes         | Yes | -            | Yes                 | -                               |
| 20                      | m00108570                     | hypothetical protein                    | -             | -      | Yes          | Yes     | -           | Yes | -            | -                   | -                               |
| 21                      | m00108600                     | TonB-dependent transporter              | -             | -      | -            | Yes     | Yes         | Yes | -            | Yes                 | -                               |
| 22                      | m00109610                     | TonB-dependent transporter              | -             | -      | -            | Yes     | Yes         | Yes | -            | -                   | -                               |
| 23                      | m00109620                     | phosphoanhydride phosphohydrolase       | <i>appA</i>   | -      | Yes          | Yes     | Yes         | Yes | -            | Yes                 | -                               |
| 24                      | m00109670                     | hypothetical protein                    | -             | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 25                      | m00109990                     | bis(5'nucleosyl)-tetraphosphatase       | <i>apaH</i>   | -      | -            | Yes     | -           | -   | -            | -                   | -                               |
| 26                      | m00110970                     | peptidyl-tRNA hydrolase                 | -             | -      | Yes          | Yes     | -           | Yes | -            | -                   | Yes                             |
| 27                      | m00111160                     | 50S ribosomal protein                   | <i>rpIW</i>   | -      | Yes          | -       | -           | Yes | -            | -                   | -                               |
| 28                      | m00112070                     | two-component system sensor protein     | <i>phoR</i>   | -      | -            | Yes     | -           | -   | -            | -                   | -                               |
| 29                      | m00113470                     | xanthine dehydrogenase subunit          | -             | -      | Yes          | Yes     | Yes         | Yes | -            | Yes                 | -                               |
| 30                      | m00113490                     | xanthine dehydrogenase subunit          | -             | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 31                      | m00114150                     | membrane protein                        | -             | -      | Yes          | -       | -           | Yes | -            | -                   | Yes                             |
| 32                      | m00114680                     | two component system sensor protein     | -             | -      | -            | Yes     | Yes         | Yes | -            | -                   | -                               |
| 33                      | m00116790                     | NAD/FAD binding protein                 | -             | -      | Yes          | -       | -           | Yes | -            | -                   | -                               |

**Table 3** Overview of the 115 genes putatively involved in bean specificity (Continued)

| Identifier <sup>a</sup> | Accession number <sup>b</sup> | Predicted function                       | Gene         | Gained | Monophyletic | 24-mers | Recombinant | HGT | Atypical GC% | Nonsynonymous sites | Aritua et al. [17] <sup>c</sup> |
|-------------------------|-------------------------------|------------------------------------------|--------------|--------|--------------|---------|-------------|-----|--------------|---------------------|---------------------------------|
| 34                      | m00117590                     | acetyltransferase                        | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 35                      | m00117600                     | transcriptional regulator                | -            | -      | Yes          | Yes     | Yes         | Yes | -            | Yes                 | -                               |
| 36                      | m00117610                     | hypothetical protein                     | -            | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 37                      | m00118890                     | hypothetical protein                     | -            | -      | Yes          | -       | -           | Yes | -            | -                   | -                               |
| 38                      | m00119700                     | putative secreted protein                | -            | -      | Yes          | -       | -           | -   | -            | -                   | -                               |
| 39                      | m00119970                     | phosphomethylpyrimidine kinase           | <i>thiD</i>  | -      | -            | Yes     | -           | Yes | -            | -                   | Yes                             |
| 40                      | m00121810                     | threonine synthase                       | <i>thrC</i>  | -      | Yes          | Yes     | -           | Yes | -            | Yes                 | -                               |
| 41                      | m00121850                     | histidyl tRNA synthetase                 | <i>hisS</i>  | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 42                      | m00121900                     | histidinol phosphatase                   | <i>hisB</i>  | -      | -            | Yes     | -           | Yes | -            | -                   | Yes                             |
| 43                      | m00123210                     | recombination factor protein             | <i>rarA</i>  | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 44                      | m00126530                     | ethanolamine permease                    | <i>eutP</i>  | -      | Yes          | Yes     | -           | Yes | -            | -                   | -                               |
| 45                      | m00126540                     | ethanolamine ammonia-lyase               | <i>eutB</i>  | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 46                      | m00127140                     | membrane protein                         | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 47                      | m00127150                     | cGMP specific phosphodiesterase          | -            | -      | Yes          | Yes     | -           | Yes | -            | Yes                 | -                               |
| 48                      | m00127220                     | hypothetical protein                     | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 49                      | m00127250                     | ubiquinol-cytochrome c reductase subunit | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 50                      | m00127260                     | ubiquinol-cytochrome c reductase subunit | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 51                      | m00127850                     | 1-phosphofructokinase                    | <i>fruK</i>  | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 52                      | m00127860                     | PTS fructose porter IIBC component       | <i>fruA</i>  | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 53                      | m00130650                     | RNA polymerase ECF-type sigma factor     | <i>rpof2</i> | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 54                      | m00130660                     | hypothetical protein                     | -            | -      | Yes          | Yes     | -           | Yes | -            | Yes                 | -                               |
| 55                      | m00130670                     | RNA-binding protein                      | -            | -      | -            | Yes     | -           | Yes | -            | -                   | Yes                             |
| 56                      | m00130690                     | inner membrane protein                   | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 57                      | m00132430                     | DNA topoisomerase IA                     | -            | -      | -            | Yes     | -           | -   | -            | Yes                 | -                               |
| 58                      | m00135840                     | membrane fusion protein                  | -            | -      | -            | Yes     | -           | -   | -            | -                   | -                               |
| 59                      | m00135870                     | sulfite reductase (NADPH) subunit        | <i>cysJ</i>  | -      | -            | Yes     | Yes         | Yes | -            | -                   | -                               |
| 60                      | m00135880                     | sulfite reductase (NADPH) subunit        | <i>cysI</i>  | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 61                      | m00135970                     | transcriptional regulator protein        | <i>cysB</i>  | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 62                      | m00135980                     | siroheme synthase                        | <i>cysG</i>  | -      | Yes          | Yes     | -           | Yes | -            | -                   | -                               |
| 63                      | m00136050                     | phosphoglycerate kinase                  | <i>pgk</i>   | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 64                      | m00136550                     | hypothetical protein                     | -            | -      | -            | Yes     | -           | -   | -            | -                   | -                               |
| 65                      | m00138140                     | two-component system sensor protein      | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 66                      | m00139330                     | dihydrolipoamide acetyltransferase       | <i>aceF</i>  | -      | -            | Yes     | Yes         | Yes | -            | -                   | -                               |

**Table 3** Overview of the 115 genes putatively involved in bean specificity (Continued)

| Identifier <sup>a</sup> | Accession number <sup>b</sup> | Predicted function                         | Gene         | Gained | Monophyletic | 24-mers | Recombinant | HGT | Atypical GC% | Nonsynonymous sites | Aritua et al. [17] <sup>c</sup> |
|-------------------------|-------------------------------|--------------------------------------------|--------------|--------|--------------|---------|-------------|-----|--------------|---------------------|---------------------------------|
| 67                      | m00139350                     | DNA glycosylase                            | -            | -      | Yes          | Yes     | -           | Yes | -            | -                   | -                               |
| 68                      | m00139420                     | membrane protein                           | -            | -      | -            | Yes     | -           | Yes | -            | Yes                 | Yes                             |
| 69                      | m00139860                     | putative secreted protein                  | -            | -      | -            | Yes     | -           | -   | -            | -                   | -                               |
| 70                      | m00139930                     | flavodoxin protein                         | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 71                      | m00140860                     | peptidoglycan binding protein              | <i>lysM</i>  | -      | Yes          | -       | -           | Yes | -            | -                   | -                               |
| 72                      | m00140880                     | hypothetical protein                       | -            | -      | -            | Yes     | -           | -   | -            | -                   | -                               |
| 73                      | m00141310                     | PhnB like protein                          | -            | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 74                      | m00141340                     | hypothetical protein                       | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 75                      | m00141400                     | hypothetical protein                       | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 76                      | m00141430                     | transmembrane protein                      | -            | -      | -            | Yes     | -           | Yes | -            | -                   | Yes                             |
| 77                      | m00141440                     | hypothetical protein                       | -            | -      | Yes          | Yes     | -           | Yes | -            | -                   | -                               |
| 78                      | m00141490                     | dipeptide epimerase                        | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 79                      | m00142140                     | putative glucosyltransferase               | -            | -      | -            | Yes     | Yes         | Yes | -            | -                   | -                               |
| 80                      | m00144120                     | type VI secretion system protein           | <i>icmF</i>  | -      | -            | Yes     | Yes         | Yes | -            | Yes                 | -                               |
| 81                      | m00144240                     | type VI secretion system kinase protein    | <i>tagE</i>  | -      | -            | Yes     | Yes         | Yes | -            | -                   | -                               |
| 82                      | m00144270                     | transmembrane sensor protein               | <i>fecR</i>  | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 83                      | m00144420                     | type VI secretion system virulence protein | <i>impE</i>  | -      | -            | Yes     | Yes         | Yes | -            | -                   | -                               |
| 84                      | m00145330                     | type III effector                          | <i>xopAD</i> | -      | -            | Yes     | Yes         | Yes | -            | Yes                 | -                               |
| 85                      | m00146940                     | membrane protein                           | -            | -      | -            | Yes     | Yes         | Yes | -            | -                   | -                               |
| 86                      | m00146970                     | LysR family transcriptional regulator      | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 87                      | m00146980                     | short chain dehydrogenase                  | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 88                      | m00146990                     | RNA 2'-phosphotransferase family protein   | -            | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 89                      | m00147020                     | hypothetical protein                       | -            | -      | Yes          | Yes     | -           | Yes | -            | -                   | -                               |
| 90                      | m00147060                     | esterase family protein                    | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 91                      | m00147090                     | putative secreted protein                  | -            | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 92                      | m00147100                     | putative secreted protein                  | -            | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 93                      | m00147120                     | ribosomal pseudouridine synthase           | -            | -      | Yes          | Yes     | -           | Yes | -            | Yes                 | -                               |
| 94                      | m00147130                     | hypothetical protein                       | -            | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 95                      | m00147140                     | putative secreted protein                  | -            | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 96                      | m00200020                     | plasmid partitioning protein               | -            | -      | Yes          | Yes     | -           | na  | -            | Yes                 | -                               |
| 97                      | m00200060                     | hypothetical protein                       | -            | Yes    | -            | Yes     | -           | na  | -            | Yes                 | -                               |
| 98                      | m00200080                     | hypothetical protein                       | -            | Yes    | -            | Yes     | -           | na  | -            | Yes                 | -                               |
| 99                      | m00200130                     | hypothetical protein                       | -            | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |

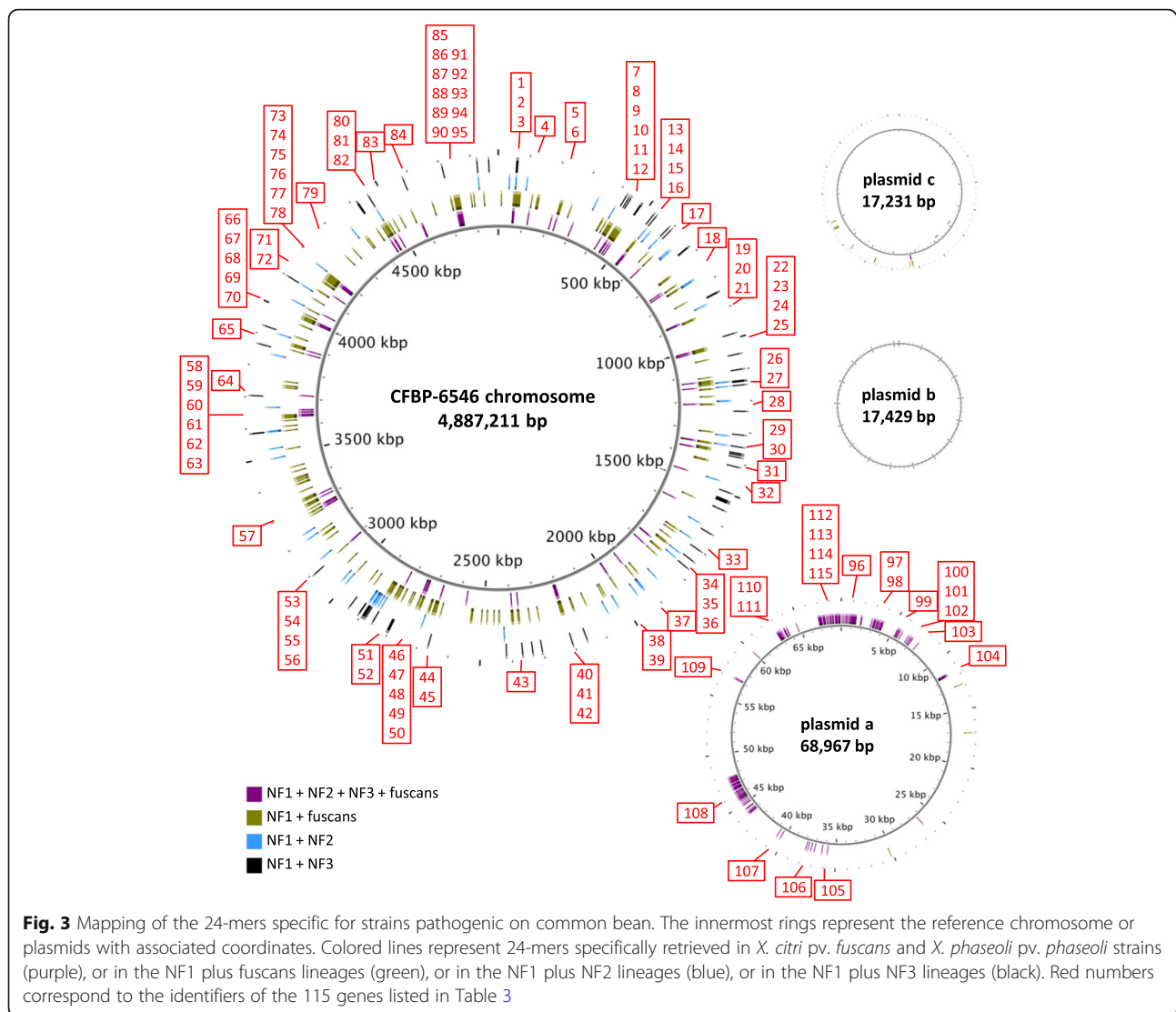
**Table 3** Overview of the 115 genes putatively involved in bean specificity (Continued)

| Identifier <sup>a</sup> | Accession number <sup>b</sup> | Predicted function                | Gene        | Gained | Monophyletic | 24-mers | Recombinant | HGT | Atypical GC% | Nonsynonymous sites | Aritua et al. [17] <sup>c</sup> |
|-------------------------|-------------------------------|-----------------------------------|-------------|--------|--------------|---------|-------------|-----|--------------|---------------------|---------------------------------|
| 100                     | m00200150                     | hypothetical protein              | -           | Yes    | -            | Yes     | -           | na  | -            | Yes                 | -                               |
| 101                     | m00200160                     | hypothetical protein              | -           | -      | -            | Yes     | -           | na  | Yes          | Yes                 | -                               |
| 102                     | m00200170                     | hypothetical protein              | -           | Yes    | -            | -       | -           | na  | -            | Yes                 | -                               |
| 103                     | m00200190                     | hypothetical protein              | -           | -      | -            | Yes     | -           | na  | -            | Yes                 | -                               |
| 104                     | m00200260                     | hypothetical protein              | -           | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 105                     | m00200790                     | transposase                       | <i>tnpA</i> | -      | -            | Yes     | -           | Yes | -            | Yes                 | -                               |
| 106                     | m00200820                     | resolvase                         | <i>tnpR</i> | -      | -            | Yes     | -           | Yes | -            | -                   | -                               |
| 107                     | m00200880                     | DNA topoisomerase IA              | -           | -      | -            | Yes     | -           | na  | -            | Yes                 | -                               |
| 108                     | m00200980                     | type IV secretion system protein  | <i>trbI</i> | Yes    | Yes          | Yes     | -           | na  | -            | Yes                 | -                               |
| 109                     | m00201120                     | membrane protein                  | -           | -      | -            | Yes     | -           | -   | -            | -                   | -                               |
| 110                     | m00201220                     | type IV conjugal transfer protein | <i>traG</i> | -      | -            | Yes     | -           | na  | -            | Yes                 | -                               |
| 111                     | m00201230                     | hypothetical protein              | -           | -      | -            | Yes     | -           | na  | -            | Yes                 | -                               |
| 112                     | m00201250                     | lytic transglycolase              | -           | Yes    | Yes          | Yes     | -           | na  | -            | Yes                 | -                               |
| 113                     | m00201260                     | type IV conjugal transfer protein | <i>traF</i> | Yes    | -            | Yes     | -           | na  | -            | Yes                 | -                               |
| 114                     | m00201270                     | hypothetical protein              | -           | Yes    | Yes          | Yes     | -           | na  | -            | Yes                 | -                               |
| 115                     | m00201290                     | hypothetical protein              | -           | Yes    | Yes          | Yes     | -           | na  | -            | Yes                 | -                               |
| Total                   | 115                           | -                                 | -           | 9      | 28           | 108     | 18          | 88  | 2            | 44                  | 9                               |

- not found in corresponding analysis

na not applicable due to the absence of corresponding CDS in outgroup

<sup>a</sup>identifiers corresponding to the numbers shown in Fig. 3<sup>b</sup>accession numbers for strain CFBP6456<sup>c</sup>genes with 100% identity over 95% of their length according to Aritua et al. [17]



by the four genetic lineages of strains responsible for CBB (in purple in Fig. 3), indicating that these regions have been shared between the four lineages recently enough to still have 100% identity between each other. Together, these results suggest that 24-mers retrieved in the chromosome correspond to more ancient HGT events than those retrieved in plasmid a.

Nucleotide synonymous substitution rates at silent sites (Ks) is an estimation of neutral evolution because it does not take into account the nonsynonymous sites that can be under selection pressure. Therefore, Ks can be used as an approximation of the time of divergence between genes or taxa, with higher Ks value meaning longer time of divergence between two sequences [27, 28]. For each of the 115 candidate genes found in CBB agents, we performed multiple alignments. We could not perform Ks analysis on 15 genes that were lacking outgroups (Table 3), therefore we tested only 100 out of

115 genes. Among these 100 genes, 18 were recombinants according to RDP software analysis (Table 3). For these 18 recombinants, Ks values were independently calculated on both sides of the breakpoints. We calculated pairwise Ks values for different combinations of strains including *X. citri* pv. *fuscans*, *X. phaseoli* pv. *phaseoli* and closely related strains including *X. citri* pv. *anacardii*, *X. citri* pv. *aurantifolii*, *X. phaseoli* pv. *syngonii*, *X. phaseoli* pv. *dieffenbachiae* and *X. phaseoli* pv. *manihotis* (Fig. 1, Additional file 3). We then used these Ks as relative time divergence estimations to infer if a HGT occurred between NF1, NF2, NF3 and/or fuscans lineages, as well as the direction of this HGT. For example, for gene *m00100580b* the mean Ks value between strains from the NF1 and fuscans lineages was  $7.40\text{e-}03 \pm 2.85\text{e-}10$ . This value was lower than the Ks values when comparing the fuscans lineage to its closest relatives from the NF2 or NF3 lineages ( $K_s = 4.58\text{e-}02 \pm$



1.32e-09 or 4.08e-02  $\pm$  1.54e-09, respectively), or when comparing the NF1 lineage to its closest relative *X. phaseoli* pv. *manihotis* ( $K_s$  = 5.33e-01  $\pm$  0.00). These results indicate that *m00100580b* was more similar between the NF1 and fuscans lineages than between these lineages and their closest relatives, meaning that *m00100580b* was horizontally transferred between the ancestors of the NF1 and fuscans lineages. Moreover, the  $K_s$  value between the NF1 lineage and the NF2 or NF3 lineages ( $K_s$  = 4.36e-02  $\pm$  1.04e-09 or 4.01e-02  $\pm$  0.00, respectively) was lower than between the NF1 lineage and *X. phaseoli* pv. *manihotis* ( $K_s$  = 5.33e-01  $\pm$  0.00). Therefore, *m00100580b* was closer between NF1 strains and other strains from group 9.6 than from its closest relatives, meaning that the horizontal transfer was directed from the fuscans lineage to the NF1 lineage. This analysis confirmed HGT for 88 out of 100 genes tested (Fig. 1, Table 3, Additional file 3). The vast majority of HGT was directed from *X. citri* pv. *fuscans* to *X. phaseoli* pv. *phaseoli*, while only four HGT occurred from *X. phaseoli* pv. *phaseoli* to *X. citri* pv. *fuscans*. In particular, 55 HGT events were detected from the fuscans lineage to the NF1 lineage. In addition to having been transferred between distant lineages, 16 and 1 genes were also transferred between the fuscans and the NF2 lineages, or the fuscans and the NF3 lineages, respectively (Fig. 1). Moreover, eight genes had  $K_s$  = 0.00  $\pm$  0.00 between the NF2 and NF3 lineages, and nine between the NF2, NF3 and fuscans lineages, suggesting that HGT events also occurred between these lineages (Fig. 1, Additional file 3). Together, our results show that several more or less important waves of HGT occurred between the ancestors of phylogenetically distant strains responsible for CBB.

Finally, GC content is often used as a mean to detect HGT from foreign origin [29]. Out of the 115 candidate CDS, only two CDS (m00105390 and m00200160) presented an atypical GC content ( $\alpha$  < 0.05) within the genome of strain CFBP6546 (Table 3). This result was not unexpected, as all strains from the NF1, NF2, NF3 and fuscans lineages have a similar GC content around 64% (Table 1), therefore HGT between these strains was not expected to result in a shift of GC content.

## Discussion

We performed a comparative genomics analysis to detect genes putatively involved in *Xanthomonas* specificity to common bean. For this, we generated the whole genome sequence from 17 strains representing the diversity of the four genetic lineages belonging to *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli*. We used a combination of comparative genomics approaches that led to the discovery of 115 genes bearing features unique to CBB agents. Out of these 115 genes, 108 were retrieved

using the SkIf tool based on specific 24-mer search [25]. Previous analyses based on identity percentage unveiled 63 genes sharing 100% identity over at least 95% of their length among strains from the NF1 and fuscans lineages [17]. Only nine of these genes were retrieved within our list of 115 genes (Table 3). This difference can be explained by the fact that we discarded most of the genes shared only by the NF1 and fuscans lineages and retained genes similar in all four genetic lineages of CBB agents (Table 2). On the other hand, we unveiled numerous genes that did not share 100% identity among the NF1 and fuscans lineages for their whole length, but instead shared small specific sequences of 24 nucleotides or more. Whether these similarities lie within functionally important domains of the encoded proteins remains to be studied.

$K_s$  comparisons, showed that a majority of these genes were involved in HGT between *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli*. Therefore, HGT was the predominant force leading to similarities between the genomes of *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli*. Finding HGT events within these genes validated our approach. In particular, SkIf was an interesting tool because in addition to being more sensitive than gene gain and loss or phylogenetic approaches, it was not based on gene alignments, thus less sensitive to annotation and/or sequencing biases. HGT events occurred at different moments of the evolution of *Xanthomonas* strains having common bean as host. The vast majority of these HGT were directed from *X. citri* pv. *fuscans* to *X. phaseoli* pv. *phaseoli*. This strongly suggest that *X. citri* pv. *fuscans* was originally pathogenic on common bean, and that *X. phaseoli* pv. *phaseoli* subsequently acquired the ability to cause CBB on bean due to successive acquisitions of novel genes and/or novel alleles coming from the three *X. citri* pv. *fuscans* lineages. This can be compared to our knowledge on the origin of the lineages and their genetic diversity. The causal agent of CBB was first isolated and identified by Smith in 1897 [30] as a yellow pigmented strain, later shown as belonging to the NF1 lineage. Burkholder later isolated the first fuscous strains from beans grown in Switzerland in 1924 [31]. However there are no data to document a putative pre-existence of one, the other, or both types of strains prior to their first identifications. The genetic diversity of the yellow and fuscous strains was revealed by various methods. Amplified or restriction fragment length polymorphism analyses [16, 32–34], amplified polymorphic DNA fragments [32, 35], pulse field gel electrophoresis [33], and multilocus sequence analysis [23] all revealed that both types of strains are more or less equivalent in terms of genetic diversity. This suggests that diversification of both lineages occurred around the same time, and thus that the ancestors of these two lineages may

have coexisted. As a consequence, *X. citri* pv. *fuscans* may be the descendant of the original CBB agent that had transferred determinants useful for adaptation on common bean to the ancestor of *X. phaseoli* pv. *phaseoli*. Therefore, the ancestor of *X. phaseoli* pv. *phaseoli* appears as a recombinant that emerged as a new common bean pathogen through the acquisition of novel genes and alleles. In quarantine areas such as Europe, Turkey, Barhain, Azerbaijan and Israel, seed lots are routinely tested using a method from the International Seed Testing Association involving isolation of bacterial strains, pathogenicity tests, and specific PCR assays [36, 37]. Our results could serve for improving PCR-based monitoring of CBB agents by designing PCR primers on genes presenting sequences unique to CBB agents and potentially important for common bean specificity. Such primers could potentially detect novel HGT of these genes in strains unrelated to *X. phaseoli* pv. *phaseoli* and *X. citri* pv. *fuscans*, thus allowing to forecast new threats potentially dangerous for common bean production.

Very diverse functions were retrieved among the proteins encoded by the 115 candidate genes. Interestingly, 44 genes contained nonsynonymous polymorphisms specific for strains responsible for CBB, suggesting that they may play an important role in common bean specificity. Although 17 of these 44 genes encoded hypothetical proteins, it appears that most other genes encoded proteins involved in pathogenicity or in the interaction with the plant environment. The type IV secretion system was particularly represented with genes encoding *TraF*, *TraG* and *TraI*, and is involved in the translocation of macromolecules such as proteins important for pathogenicity, or DNA for mediating HGT [38, 39]. Thus, sharing similar proteins of the type IV secretion system could favour HGT among strains responsible for CBB. Indeed, strains from the NF1, NF2, NF3 and *fuscans* lineages have been found in La Réunion Island in 2000 (Table 1), indicating that sympatry exists among all these lineages, rendering further HGT events possible [18]. Three genes encoding proteins related to the type III secretion system were retrieved. XopA [40] and HpaH [41] are two proteins that may be involved in the structure of the type III secretion system, while XopAD is a type III effector of unknown function consisting of multiple semi-conserved 42 amino acids SKW repeats [42, 43]. The type III secretion system is pivotal for the virulence of most Gram negative plant pathogenic bacteria, and repertoires of type III effectors have been described as potentially important factors for host specificity and host adaptation in *Xanthomonas* [7, 44, 45] and other genera such as *Pseudomonas* [46]. Moreover, our analysis pointed out one diguanilate cyclase and one cGMP specific phosphodiesterase, two proteins involved in the metabolism of cyclic di-GMP that may play a role in

biofilm formation [47] and pathogenicity [48]. One TonB-dependent transporter was also retrieved. TonB-dependent transporters are outer membrane receptors involved in molecule uptake such as iron siderophore complexes or nutrients and may play a role in host specificity [3, 49]. Other proteins putatively involved in pathogenicity were retrieved, such as ThrC, a threonine synthase involved in the virulence of *X. oryzae* pv. *oryzicola* in rice [50], XpsD, an outer membrane protein from the type II secretion system that is putatively involved in the secretion of cell wall degradative enzymes during infection [51], or IcmF, a protein of the type VI secretion system, which is involved in the interaction with other bacteria and may participate in pathogenicity [52, 53]. One flavine reductase and one xanthine dehydrogenase, two proteins putatively involved in oxydoreduction pathways were retrieved, and may be involved in the response to stress during the interaction with common bean. In addition to genes putatively involved in pathogenicity or virulence, our analysis unveiled genes involved in more general metabolism pathways, such as PhnB involved in the biosynthesis of tryptophan [54], or two DNA topoisomerases involved in the relaxation of the supercoiled DNA molecule during transcription, replication or recombination [55]. On one hand, our analysis unveiled only one or few genes within a given function, while the functions retrieved correspond to pathways often involving dozens of genes. This suggested that slight modifications within a given pathway would be sufficient to impact host specificity. On the other hand, the genes retrieved here encompass almost all the stages of host plant colonization by the bacteria, from the ability to mobilize trophic resources for multiplication to the interaction with other microorganisms, biofilm formation, response to oxidative stress, and inhibition of plant defences. Therefore, the ability to infect a particular plant seems to require not just one or a few adaptative determinants but an arsenal of factors allowing a global adaptation to a specific niche including the plant and, as a consequence a fine tuning and coordination of the activity of these determinants.

Interestingly, 19 of these 44 candidate genes were retrieved on plasmid a, suggesting that this plasmid played a major role for pathological convergence of CBB agents. Plasmid a carries an additional type III effector gene encoding an effector from the Transcription Activator-Like (TAL) family that was horizontally transferred between the NF1, NF2, NF3 and *fuscans* lineages [18]. Plasmids are genetic elements that favour HGT, but transfers of whole plasmids often induce a fitness cost for the bacteria [56]. More generally, horizontally transferred genes tend to be lost if not providing selective advantages for recipient strains [57]. Interestingly, the nine candidate genes retrieved by the gain and loss approach were all

located on plasmid a (Table 3). The maintenance of these novel genes in the four genetic lineages of CBB agents is a testament to the importance of these genes for the bacteria. Except for two genes encoding proteins involved in the type IV secretion system, the other seven genes encoded proteins of unknown function. It would be interesting to perform functional characterization of these genes, and further analyse their implication in common bean specificity. Analysing the expression patterns during infection would be a natural extension of this study, and a first step towards functional validation of these genes.

## Conclusion

Together, our results indicate that consecutive waves of HGT occurred between phylogenetically distant *Xanthomonas* strains able to cause the same symptoms on the same host plant: common bean. These HGT led to specific combinations of genes only retrieved in strains responsible for CBB, which provided new insights into the evolution of these bacteria towards infecting common bean. Mining for candidate genes for host specificity could be generalized to other polyphyletic pathovars such as pathovars *euvesicatoria*, *vesicatoria*, *perforans*, and *gardneri* forming a group of strains pathogenic on pepper and tomato [8]. Such analyses could both give new information on the molecular bases of host specificity, and provide new tools for enhancing epidemiological surveillance of strains pathogenic on a given host, or detecting recombinant strains presenting a high potential of emergence through the acquisition of novel genes.

## Methods

### Bacterial genomes and strains

All strains used in this study are listed in Table 1. The strains used for genome sequencing were provided by the CIRM-CFBP (International Center for Microbial Ressources - French Collection for Plant-associated Bacteria, [https://www6.inra.fr/cirm\\_eng/](https://www6.inra.fr/cirm_eng/)). Genome sequencing was performed using the following procedure. Genomic DNA was prepared from overnight liquid cultures of bacteria previously grown on 10% TSA medium (tryptone at 1.7 g/L, soybean peptone at 0.3 g/L, glucose at 0.25 g/L, NaCl at 0.5 g/L, K<sub>2</sub>HPO<sub>4</sub> at 0.5 g/L, agar at 15 g/L, pH 7.2) for 2 days at 28 °C. DNA was extracted and purified by the method of Klotz and Zimm [58]. Illumina sequencing was performed by Genoscope (20 strains, paired end reads of 300/500 bp) or GATC Biotech (three strains, with combined paired-end reads of ca. 250 bp and 3 kb mate-pair reads). Genome assembly was performed using a combination of SOAPdenovo (version 2.04) [59], SOAPGapCloser (version 1.12) [60] and Velvet (version 1.2.02) [61] assemblers. Sequenced

genomes were estimated to be > 93% complete and < 3% contaminated (Additional file 4) using CheckM (version 1.0) [62]. The pathogenicity of all CBB strains was confirmed on common bean plants from cultivar Flavert as described in Ruh et al. [18]. The seeds from cultivar Flavert were kindly provided by Vilmorin (La Ménétré, France) and are available at the bean collection of the CIAT (Center for Tropical Agriculture, Colombia, <http://genebank.ciat.cgiar.org/genebank/main.do>).

### Annotation and phylogenomic analyses

Structural and functional annotation of whole genome assemblies was performed using the automated pipeline Eugene-PP (version 1.2) [63], using SWISS-PROT as protein database and training protein database (<http://www.uniprot.org/>). Additional functional annotation of all predicted CDS was performed with InterProScan (version 4) [64]. A presence/absence matrix of ortholog groups was constructed using OrthoMCL (version 2.0) on amino acid sequences from all predicted CDS at an inflation index of 1.5 [65]. This matrix was then used for defining core and pan genomes. Phylogenetic trees were constructed using CTree (version 4.2) [66] using the amino acid sequences of all predicted CDS from the 75 genomes used in this study. CDS gains and losses were analysed using the Most Parsimonious Reconstruction function from the APE package (version 3.2) [67] to search for the most parsimonious succession of events explaining the repertoire of ortholog groups at each node of the phylogenetic tree.

### Searching for genes monophyletic for *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* strains

A phylogenetic approach was used to search for genes for which strains from *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* form a monophyletic group. For this we selected 3202 CDS using an R script to search the orthology matrix for genes that were present in all *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* strains and in at least one strain from Rep-PCR group 9.2 or 9.4 [11] plus one another strain from Rep-PCR group 9.5 or 9.6, in order to avoid getting trees where *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* appear monophyletic due to a lack of corresponding genes in the strains in between. CDS were aligned using MAFFT (version 7) with L-INS-I strategy [68]. Neighbour-joining trees were constructed using APE (version 3.2) under the Kimura 80 model [67]. CDS monophyletic for all *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* strains, or alternatively for the NF1 and another lineage (i.e. NF2, NF3, or *fuscans*), were retrieved using the APE package (version 3.2) [67].

### Searching for genes containing *k*-mers specific for *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* strains

A *k*-mer-based approach was used to search for genes containing short specific sequences present in all strains from *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* but absent in other strains. For this, we used Sklf (version 1.0) [25] with a *k*-mer size of 24 (or 24-mer), and using *X. citri* pv. *fuscans* strain 4834-R genome as reference. The same approach was used to search for genes containing 24-mers absent in strains belonging to *X. citri* pv. *fuscans* and *X. phaseoli* pv. *phaseoli* but conserved in all other strains from the *X. axonopodis* species complex, using *X. citri* pv. *anacardii* strain CFBP2913 genome as reference.

### Recombination and HGT analyses

The 115 genes presenting specific traits of adaptation to common bean were aligned using MAFFT (version 7) with L-INS-I strategy [68]. Intragenic recombination events were then searched using a suite of programs implemented in RDP (version 4.16) [69], RDP [70], Geneconv [71], MaxChi [72], Chimaera [73], Bootscan [74] and 3seq [75]. Default parameters were used for each method except for Bootscan (window = 150, step = 20, neighbor joining trees, 200 replicates, 95% cut-off, J&N model with Ti:Tv = 2, coefficient of variation = 2). *Ks* was calculated using DNAsp (version 5) [76]. For each gene, the occurrence and dating of HGT events were estimated by comparing *Ks* values from 28 different pairwise combinations listed in Additional file 3. For example, NF1 and *fuscans* strains belong to phylogenetically distant strains, thus if the *Ks* between strains from genetic lineages NF1 and *fuscans* was lower than the mean *Ks* between other lineages, it was indicative of recent HGT between the ancestors of NF1 and *fuscans*. Direction of events were assessed by comparing the *Ks* values for outgroups belonging to Rep-PCR groups 9.4 and 9.6 (Fig. 1). For recombinants, separate analyses were performed for each region on both sides of the recombination point.

### Additional files

**Additional file 1:** CDS content per strain. (PDF 406 kb)

**Additional file 2:** OrthoMCL matrix. (XLSX 10182 kb)

**Additional file 3:** *Ks* analysis on 100 CDS presenting characteristics unique to CBB agents. (XLSX 64 kb)

**Additional file 4:** CheckM analysis of the genomes sequenced in this study. (XLSX 11 kb)

### Abbreviations

CBB: Common bacterial blight of bean; CDS: Coding sequence; HGT: Horizontal gene transfer; *Ks*: Nucleotide synonymous substitution rate at silent sites; NF1, 2 or 3: Non-fuscos lineages 1, 2 or 3; PTI: Pattern

triggered immunity (PTI); rep-PCR: Repetitive-sequence-based Polymerase Chain Reaction; TAL: Transcription activator-like

### Acknowledgements

The authors thank Sébastien Carrère for his help on genome annotations and Laurent D. Noël for sharing the genome sequence of strain CFBP6865. The authors also thank the French Network on Xanthomonads (FNX) (<https://www.reseau-xantho.org/>) for recurrent scientific exchanges on *Xanthomonas*. Authors benefited from interactions promoted by COST Action FA 1208 (<https://www.cost-sustain.org>). We thank the CIRM-CFBP (Beaucouzé, France) for strain preservation and supply.

### Funding

The research leading to these results has received grants from the Genoscope (3X 154/AP2006-2007, XANTHOMICS 18/AP2009-2010) and the French National Research Agency (XANTHOMIX ANR-2010-GENM-013-02). LSG was funded by a postdoctoral grant from the XANTHOMIX ANR project, and another postdoctoral grant from Angers-Loire Metropole, France. MR was funded by a PhD grant from Angers-Loire Metropole, France. This work was supported by France Génomique National infrastructure, funded as part of “Investissement d’avenir” program managed by Agence Nationale pour la Recherche (contrat ANR-10-INBS-0009).

### Availability of data and materials

The datasets (genome sequences) generated through various projects and used in this study have been deposited in GenBank under accession numbers listed in Table 1.

### Authors’ contributions

MAJ, LSG and NWGC designed the study. This study was partly lead by LSG during her post-doc and is part of the doctoral research of MR supervised by MAJ and NWGC. MAJ, NWGC, LG and RK obtained grants to fund the study. VB managed the genome sequencing at Genoscope. NWGC, LSG, MAJ, MB, MR, SB, AD, and VB contributed to data analyses and interpretation. NWGC, MR and MAJ wrote the manuscript with inputs from all co-authors. All authors read and approved the final version of the manuscript.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>IRHS, INRA, AGROCAMPUS OUEST, Université d’Angers, SFR4207 QUASAV, 42, rue Georges Morel, 49071 Beaucouzé, France. <sup>2</sup>CEA/DSV/IG/Genoscope, 2 rue Gaston Crémieux, BP5706, 91057 Evry, France. <sup>3</sup>CIRAD, UMR PVBM, F-97410 Saint-Pierre, La Réunion, France. <sup>4</sup>IRD, CIRAD, Université de Montpellier, IPME, Montpellier, France.

Received: 11 February 2018 Accepted: 31 July 2018

Published online: 13 August 2018

### References

- Bäumler A, Fang FC. Host specificity of bacterial pathogens. Cold Spring Harb Perspect Med. 2013;3:a010041. Available from: <http://perspectivesinmedicine.cshlp.org/content/3/12/a010041>.
- Büttner D, Bonas U. Regulation and secretion of *Xanthomonas* virulence factors. FEMS Microbiol Rev. 2010;34:107–33.
- Jacques M-A, Arlat M, Boulanger A, Boureau T, Carrère S, Cesbron S, et al. Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*: French Network on Xanthomonads (FNX). Annu Rev Phytopathol. 2016;54:163–87.
- Hayward AC. The hosts of *Xanthomonas*. In: *Xanthomonas*; 1993. p. 1–119.



5. Dye DW, Bradbury JF, Goto M, Hayward AC, Lelliott RA, Schroth MN. International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains. *Rev Plant Pathol.* 1980; 59:153–9.
6. Mhedbi-Hajri N, Darrasse A, Pigné S, Durand K, Fouteau S, Barbe V, et al. Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads. *BMC Evol Biol.* 2011;11:67. BioMed Central Ltd. Available from: <https://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-11-67>.
7. Hajri A, Brin C, Hunault G, Lardeux F, Lemaire C, Manceau C, et al. A «repertoire for repertoire» hypothesis: repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS One.* 2009;4:e6632.
8. Schwartz AR, Potnis N, Timilsina S, Wilson M, Patané J, Martins J, et al. Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. *Front Microbiol.* 2015;6:535.
9. Vauterin L, Hoste B, Kersters K, Swings J. Reclassification of *Xanthomonas*. *Int J Syst Bacteriol.* 1995;45:472–89. Available from: <http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-45-3-472>.
10. Vauterin L, Rademaker J, Swings J. Synopsis on the taxonomy of the genus *Xanthomonas*. *Phytopathology.* 2000;90:677–82.
11. Rademaker JLW, Louws FJ, Schultz MH, Rossbach U, Vauterin L, Swings J, et al. A comprehensive species to strain taxonomic framework for *Xanthomonas*. *Phytopathology.* 2005;95:1098–1111.
12. Constantin EC, Cleenwerck I, Maes M, Baeyen S, Van Malderghem C, De Vos P, et al. Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathol.* 2016;65:792–806.
13. Oren A, Garrity GM. Notification of changes in taxonomic opinion previously published outside the IJSEM. *Int J Syst Evol Microbiol.* 2017;67:2081–6. Available from: <http://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.002071>.
14. EFSA PLH Panel (EFSA Panel on Plant Health). Scientific opinion on the pest categorisation of *Spodoptera littoralis*. *EFSA J.* 2015;13:3987.
15. Belete T, Bastas K. Common Bacterial Blight (*Xanthomonas axonopodis* pv. *phaseoli*) of Beans with Special Focus on Ethiopian Condition. *J Plant Pathol Microbiol.* 2017;8:403. Available from: <https://www.omicsonline.org/open-access/common-bacterial-blight-xanthomonas-axonopodis-pv-phaseoli-of-beans-with-special-focus-on-ethiopian-condition-2157-7471-1000403.php?aid=87779>.
16. Alavi SM, Sanjari S, Durand F, Brin C, Manceau C, Poussier S. Assessment of the genetic diversity of *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans* as a basis to identify putative pathogenicity genes and a type III secretion system of the SPI-1 family by multiple suppression subtractive h. *Appl Environ Microbiol.* 2008;74:3295–301.
17. Aritua V, Harrison J, Sapp M, Buruchara R, Smith J, Studholme DJ. Genome sequencing reveals a new lineage associated with lablab bean and genetic exchange between *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans*. *Front Microbiol.* 2015;6:1–18. Available from: <http://journal.frontiersin.org/article/10.3389/fmicb.2015.01080>
18. Ruh M, Briand M, Bonneau S, Jacques M-A, Chen NWG. *Xanthomonas* adaptation to common bean is associated with horizontal transfers of genes encoding TAL effectors. *BMC Genomics.* 2017;18:670. Available from: <http://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-017-4087-6>
19. Simpson AJG, Reinach FC, Arruda P, Abreu FA, Acencio M, Alvarenga R, et al. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature.* 2000;406:151–9.
20. van Sluys MA, de Oliveira MC, Miyaki CY, Furlan LR, Camargo LEA, Silva ACR, et al. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. *J Bacteriol.* 2003;185:1018–26.
21. Taghavi S, Garafola C, Monchy S, Newman L, Hoffman A, Weyens N, Barac T, et al. Genome survey and characterization of endophytic bacteria exhibiting a beneficial effect on growth and development of poplar trees. *Appl Environ Microbiol.* 2009;75:748–57.
22. Jacques M-A, Arlat M, Boulanger A, Boureau T, Carrère S, Cesbron S, et al. Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annu Rev Phytopathol.* 2016;54:163–87. Available from: <http://www.annualreviews.org/doi/10.1146/annurev-phyto-080615-100147>.
23. Mhedbi-Hajri N, Hajri A, Boureau T, Darrasse A, Durand K, Brin C, et al. Evolutionary history of the plant pathogenic bacterium *Xanthomonas axonopodis*. *PLoS One.* 2013;8:e58474.
24. Young JM, Park DC, Shearman HM, Fargier E. A multilocus sequence analysis of the genus *Xanthomonas*. *Syst Appl Microbiol.* 2008;31:366–77.
25. Briand M, Gaborieau R, Jacques M-A, Barret M, Boureau T, Gaillard S, et al. Sklf: a tool for rapid identification of genes or regulators of interest [version 1; not peer reviewed]. *F1000Res.* 2016;5:1593. (poster)
26. Darrasse A, Bolot S, Serres-Giardi L, Charbit E, Boureau T, Saux MF-L, et al. High-quality draft genome sequences of *Xanthomonas axonopodis* pv. *glycines* strains CFBP 2526 and CFBP 7119. *Genome Announc.* 2013;1:e01036–13.
27. Ochman H, Elwyn S, Moran NA. Calibrating bacterial evolution. *Proc Natl Acad Sci U S A.* 1999;96:12638–12643. Available from: <http://www.pnas.org/content/96/22/12638>.
28. Kuo C-H, Ochman H. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biol Direct.* 2009;4:35. Available from: <https://biologydirect.biomedcentral.com/articles/10.1186/1745-6150-4-35>.
29. Ravenhall M, Nives S, Lassalle F, Dessimoz C. Inferring horizontal gene transfer; 2015. p. 1–16.
30. Zaumeyer WJ. The bacterial blight of beans caused by *Bacterium phaseoli*. Washington: United States Department of Agriculture; 1930.
31. Burkholder WH. The bacterial diseases of the bean: a comparative study. 1930.
32. Birch PRJ, Hyman LJ, Taylor R, Opio AF, Bragard C, Toth IK. RAPD PCR-based differentiation of *Xanthomonas campestris* pv. *Phaseoli* and *Xanthomonas campestris* pv. *phaseoli* var. *fuscans*. *Eur J Plant Pathol.* 1997;103:809–14.
33. Chan JWYF, Goodwin PH. Differentiation of *Xanthomonas campestris* pv. *phaseoli* from *Xanthomonas campestris* pv. *phaseoli* var. *fuscans* by PFGE and RFLP. *Eur J Plant Pathol.* 1999;105:867–78.
34. Lazo GR, Gabriel DW. Conservation of plasmid DNA sequences and pathovar identification of strains of *Xanthomonas campestris*. *Phytopathology.* 1987;77:448–53.
35. Mkandawire ABC, Mabagala RB, Guzmán P, Gepts P, Gilbertson RL. Genetic diversity and pathogenic variation of common blight bacteria (*Xanthomonas campestris* pv. *phaseoli* and *X. campestris* pv. *phaseoli* var. *fuscans*) suggests pathogen coevolution with the common bean. *Phytopathology.* 2004;94:593–603.
36. Audy P, Laroche A, Saindon G, Huang HC, Gilbertson RL. Detection of the bean common blight bacteria, *Xanthomonas campestris* pv. *phaseoli* and *X. c. phaseoli* var. *fuscans*, using the polymerase chain reaction. *Mol Plant Pathol.* 1994;4:1185–92.
37. Grimault V, Olivier V, Rolland M, Darrasse A, Jacques MA. Seed health testing methods. 7-021: Detection of *Xanthomonas axonopodis* pv. *phaseoli* on *Phaseolus vulgaris*. ISTA Int rules seed testing Annex to Chapter 7 Seed Health methods 7-021 Int Seed Test Assoc Baserdorf, Switzerland, 1–20. 2014.
38. Fronzes R, Christie PJ, Waksman G. The structural biology of type IV secretion systems. *Nat Rev Microbiol.* 2009;7:703–14.
39. Wallden K, Rivera-Calzada A, Waksman G. Type IV secretion systems: versatility and diversity in function. *Cell Microbiol.* 2010;12:1203–12.
40. Noël L, Thieme F, Nennstiel D, Bonas U. Two novel type III-secreted proteins of *Xanthomonas campestris* pv. *vesicatoria* are encoded within the hrp pathogenicity island. *J Bacteriol.* 2002;184:1340–8.
41. Hausner J, Hartmann N, Jordan M, Büttner D. The predicted lytic transglycosylase HpaH from *xanthomonas campestris* pv. *vesicatoria* associates with the Type III secretion system and promotes effector protein translocation. *Infect Immun.* 2017;85:1–18.
42. Teper D, Burstein D, Salomon D, Gershovitz M, Pupko T, Sessa G. Identification of novel *Xanthomonas euvesicatoria* type III effector proteins by a machine-learning approach. *Mol Plant Pathol.* 2016;17:398–411.
43. Escalon A, Javegny S, Vernière C, Noël LD, Vital K, Poussier S, et al. Variations in type III effector repertoires, pathological phenotypes and host range of *Xanthomonas citri* pv. *citri* pathotypes. *Mol Plant Pathol.* 2013;14:483–96.
44. Hersemann L, Wibberg D, Blom J, Goesmann A, Widmer F, Vorholter F-J, et al. Comparative genomics of host adaptive traits in *Xanthomonas translucens* pv. *graminis*. *BMC Genomics.* 2017;18:35. Available from: <https://doi.org/10.1186/s12864-016-3422-7>.
45. Barak JD, Vancheva T, Lefevre P, Jones JB, Timilsina S, Minsavage GV, et al. Whole-genome sequences of *Xanthomonas euvesicatoria* strains clarify taxonomy and reveal a stepwise erosion of type 3 effectors. *Front Plant Sci.* 2016;7:1805. Available from: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01805/full>.
46. Baltrus DA, Nishimura MT, Romanchuk A, Chang JH, Mukhtar MS, Cherkis K, et al. Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLoS Pathog.* 2011;7:e1002132.

47. Newell PD, Boyd CD, Sondermann H, O'Toole GA. A c-di-GMP effector system controls cell adhesion by inside-out signaling and surface protein cleavage. *PLoS Biol.* 2011;9:e1000587.
48. Su J, Zou X, Huang L, Bai T, Liu S, Yuan M, et al. DgcA, a diguanylate cyclase from *Xanthomonas oryzae* pv. *oryzae* regulates bacterial pathogenicity on rice. *Sci Rep.* 2016;6:25978.
49. Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denancé N, et al. Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. *PLoS One.* 2007;2:e224.
50. Guo W, Cui YP, Li YR, Che YZ, Yuan L, Zou LF, et al. Identification of seven *Xanthomonas oryzae* pv. *oryzicola* genes potentially involved in pathogenesis in rice. *Microbiology.* 2012;158:505–18.
51. Solé M, Scheibner F, Hoffmeister AK, Hartmann N, Hause G, Rother A, et al. *Xanthomonas campestris* pv. *vesicatoria* secretes proteases and xylanases via the Xps type II secretion system and outer membrane vesicles. *J Bacteriol.* 2015;197:2879–93.
52. Records AR, Gross DC. Sensor kinases RetS and LadS regulate *Pseudomonas syringae* type VI secretion and virulence factors. *J Bacteriol.* 2010;192:3584–96.
53. Alteri CJ, Mobley HLT. The versatile type VI secretion system. *Microbiol Spectr.* 2016;4:1–26.
54. Xie G, Bonner CA, Brettin T, Gottardo R, Keyhani NO, Jensen RA. Lateral gene transfer and ancient paralogy of operons containing redundant copies of tryptophan-pathway genes in *Xylella* species and in heterocystous cyanobacteria. *Genome Biol.* 2003;4:R14.
55. Hashimi H, Ziková A, Panigrahi AK, Stuart KD, Lukeš J. TbrGG1, an essential protein involved in kinetoplast RNA metabolism that is associated with a novel multiprotein complex. *RNA.* 2008;14:970–80.
56. Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. *Nat Rev Genet.* 2015;16:472–82. Nature Publishing Group. Available from: <http://www.nature.com/doi/10.1038/nrg3962>.
57. Koskiniemi S, Sun S, Berg OG, Andersson DI. Selection-driven gene loss in bacteria. *PLoS Genet.* 2012;8:1–7. Available from: <http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002787>.
58. Klotz LC, Zimm BH. Size of DNA determined by viscoelastic measurements: results on bacteriophages, *Bacillus subtilis* and *Escherichia coli*. *J Mol Biol.* 1972;72:779–800.
59. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 2010;20:265–72. Available from: <http://genome.cshlp.org/content/20/2/265.short>.
60. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience.* 2012;1:18.
61. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008;18:821–9.
62. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. 2015;25:1043–55.
63. Sallet E, Gouzy J, Schiex T. EuGene-PP: a next-generation automated annotation pipeline for prokaryotic genomes. *Bioinformatics.* 2014;30:2659–61.
64. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. *Nucleic Acids Res.* 2005;33:116–20.
65. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003;13:2178–89. Available from: <http://www.genome.org/cgi/doi/10.1101/gr.1224503>.
66. Qi J, Luo H, Hao B. CVTree: a phylogenetic tree reconstruction tool based on whole genomes. *Nucleic Acids Res.* 2004;32:45–7.
67. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics.* 2004;20:289–90.
68. Katoh K, Kuma KI, Toh H, Miyata T. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 2005;33:511–8. Available from: <https://academic.oup.com/nar/article/33/2/511/2549118>. [cited 21 July 2014]
69. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 2015;1:1–5. Available from: <http://ve.oxfordjournals.org/cgi/doi/10.1093/ve/vev003>.
70. Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics.* 2000;16:562–3.
71. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. *Virology.* 1999;265:218–25. Available from: <https://www.sciencedirect.com/science/article/pii/S0042682299005697via%3Dihub>.
72. Smith JM. Analyzing the mosaic structure of genes. *J Mol Evol.* 1992;34:126–9.
73. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci.* 2001;98:13757–62.
74. Martin DP, Posada D, Crandall KA, Williamson C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retrovir.* 2005;21:98–102.
75. Boni MF, Posada D, Feldman MW. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics.* 2007;176:1035–47.
76. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009;25:1451–2. Available from: <https://academic.oup.com/bioinformatics/article/25/11/1451/332507>. [cited 11 July 2014]
77. Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, et al. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genomics.* 2009;15:1–15.
78. Roux B, Bolot S, Guy E, Denancé N, Lautier M, Jardinaud M-F, et al. Genomics and transcriptomics of *Xanthomonas campestris* species challenge the concept of core type III effectome. *BMC Genomics* [Internet]. *BMC Genomics.* 2015;16:975. Available from: <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-015-2190-0>.
79. Qian W, Jia Y, Ren S, He Y, Feng J, Lu L, et al. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome Res.* 2005;757–67.
80. da Silva ACR, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, et al. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature.* 2002;417:459–63.
81. Vorhölter F-J, Schneiker S, Goesmann A, Krause L, Bekel T, Kaiser O, et al. The genome of *Xanthomonas campestris* pv. *campestris* B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. *J Biotechnol.* 2008;134:33–45.
82. Studholme DJ, Kemen E, MacLean D, Schornack S, Aritua V, Thwaites R, et al. Genome-wide sequencing data reveals virulence factors implicated in banana *Xanthomonas* wilt. *FEMS Microbiol Lett.* 2018;2:182–92.
83. Bogdanove AJ, Koebnik R, Lu H, Furutani A, Angioli S V, Patil PB, et al. Two New Complete Genome Sequences Offer Insight into Host and Tissue Specificity of Plant Pathogenic *Xanthomonas* spp. *J Bacteriol.* 2011;193:5450–64.
84. Bolot S, Guy E, Carrere S, Barbe V, Arlat M, Noël LD. Genome Sequence of *Xanthomonas campestris* pv. *campestris* Strain Xca5. *Genome Announc* [Internet]. 2013;1:32–3. Available from: <http://genomea.asm.org/cgi/doi/10.1128/genomeA.00032-12>.
85. Moreira LM, Almeida NF, Potnis N, Digiampietri LA, Adi SS, Bortolossi JC, et al. Novel insights into the genomic basis of citrus canker based on the genome sequences of two strains of *Xanthomonas fuscans* subsp. *aurantifolii*. *BMC Genomics.* 2010;11.
86. Gordon JL, Lefeuve P, Escalon A, Barbe V, Cruveiller S, Gagnevin L, et al. Comparative genomics of 43 strains of *Xanthomonas citri* pv. *citri* reveals the evolutionary events giving rise to pathotypes with different host ranges. *BMC Genomics* [Internet]. *BMC Genomics.* 2015;16. Available from: <https://doi.org/10.1186/s12864-015-2310-x>.
87. Cunnac S, Bolot S, Serna F, Ortiz E, Szurek B, Noël LD, et al. High-Quality Draft Genome Sequences of Two *Xanthomonas citri* pv. *malvacearum* Strains. *Genome Announc.* 2013;1:4–5.
88. Midha S, Ranjan M, Sharma V, Pinnaka AK, Patil PB. Genome sequence of *xanthomonas citri* pv. *Mangiferaeindicae* strain LMG 941. *J Bacteriol.* 2012;194:3031.
89. Gagnevin L, Bolot S, Gordon JL, Pruvost O, Vernière C, Robène I, et al. Draft Genome Sequence of *Xanthomonas axonopodis* pv. *allii* Strain CFBP 6369. *Genome Announc* [Internet]. 2014;2:e00727-14. Available from: <http://genomea.asm.org/content/2/4/e00727-14.full>.
90. Jalan N, Aritua V, Kumar D, Yu F, Jones JB, Graham JH, et al. Comparative genomic analysis of *Xanthomonas axonopodis* pv. *citrumelo* F1, which causes citrus bacterial spot disease, and related strains provides insights into virulence and host specificity. *J Bacteriol.* 2011;193:6342–57.
91. Thieme F, Koebnik R, Bekel T, Berger C, Boch J, Büttner D, et al. Insights into Genome Plasticity and Pathogenicity of the Plant Pathogenic Bacterium *Xanthomonas campestris* pv. *vesicatoria* Revealed by the Complete Genome Sequence Insights into Genome Plasticity and Pathogenicity of the Plant Pathogenic Bacterium. *J Bacteriol.* 2005;187:7254–66.

92. Potnis N, Krasileva K, Chow V, Almeida NF, Patil PB, Ryan RP, et al. Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genomics* [Internet]. BioMed Central Ltd; 2011;12:146. Available from: <http://www.biomedcentral.com/1471-2164/12/146>.
93. Kimbrel JA, Givan SA, Temple TN, Johnson KB, Chang JH. Genome sequencing and comparative analysis of the carrot bacterial blight pathogen, *Xanthomonas hortorum* pv. *carotae* M081, for insights into pathogenicity and applications in molecular diagnostics. *Mol Plant Pathol*. 2011;12:580–94.
94. Lee BM, Park YJ, Park DS, Kang HW, Kim JG, Song ES, et al. The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. *Nucleic Acids Res*. 2005;33:577–86.
95. Salzberg SL, Sommer DD, Schatz MC, Phillippy AM, Rabinowicz PD, Tsuge S, et al. Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. *BMC Genomics*. 2008;9:1–16.
96. Robène I, Bolot S, Pruvost O, Arlat M, Noël LD, Carrère S, et al. High-Quality Draft Genome Sequences of Two *Xanthomonas* Pathotype Strains Infecting Aroid Plants. 2016;4:5–6.
97. Rodriguez-R LM, Grajales A, Arrieta-Ortiz ML, Salazar C, Restrepo S, Bernal A. Genomes-based phylogeny of the genus *Xanthomonas*. *BMC Microbiol* [Internet]. BioMed Central Ltd; 2012;12:43. Available from: <https://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-12-43>.
98. Studholme DJ, Wasukira A, Paszkiewicz K, Aritua V, Thwaites R, Smith J, et al. Draft genome sequences of *xanthomonas sacchari* and two banana-associated *xanthomonads* reveal insights into the *xanthomonas* group 1 clade. *Genes (Basel)*. 2011;2:1050–65.
99. Pesce C, Bolot S, Cunnac S, Portier P, Fisher-Le Saux M, Jacques M-A, et al. High-Quality Draft Genome Sequence of the *Xanthomonas translucens* pv. *cerealis* Pathotype Strain CFBP 2541. *Genome Announc*. 2015a;3:1–2.
100. Pesce C, Bolot S, Berthelot E, Bragard C, Cunnac S, Fischer-Le Saux M, et al. Draft Genome Sequence of *Xanthomonas translucens* pv. *graminis* Pathotype Strain CFBP 2053. *Genome Announc* [Internet]. 2015b;3:15–7. Available from: <https://mra.asm.org/content/3/5/e01174-15>.
101. Wichmann F, Vorhölter FJ, Hersemann L, Widmer F, Blom J, Niehaus K, et al. The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. *Mol Plant Pathol*. 2013;14:576–88.
102. van Sluys MA, de Oliveira MC, Monteiro-Vitorello CB, Miyaki CY, Furlan LR, Camargo LEA, et al. Comparative Analyses of the Complete Genome Sequences of Pierce 's Disease and Citrus Variegated Chlorosis Strains of *Xylella fastidiosa*. *J Bacteriol* [Internet]. 2003;185:1018–26. Available from: <http://jb.asm.org/cgi/doi/10.1128/JB.185.3.1018-1026.2003>.
103. Rademaker JLW, Hoste B, Louws FJ, Kersters K, Swings J, Vauterin L, et al. Comparison of AFLP and rep-PCR genomic fingerprinting with DNA – DNA homology studies: *Xanthomonas* as a model system. *Int J Syst Evol Microbiol*. 2000;50:665–77.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

