Deep phylo-taxono genomics reveals *Xylella* as a variant lineage of plant associated *Xanthomonas* with *Stenotrophomonas* and *Pseudoxanthomonas* as misclassified relatives

Kanika Bansal[1^], Sanjeet Kumar[1^], Amandeep Kaur[1], Anu Singh[1], Prabhu B. Patil[1*]

[*1]Bacterial Genomics and Evolution Laboratory, CSIR-Institute of Microbial Technology, Chandigarh.

^Equal Contribution

*Corresponding author

Prabhu B. Patil

Email: pbpatil@imtech.res.in

Keywords

Xanthomonadaceae, plant pathogen, variant-lineage, whole-genome sequencing, comparative genomics

Abstract

Genus *Xanthomonas* is a group of phytopathogens which is phylogenetically related *to Xylella*, *Stenotrophomonas* and *Pseudoxanthomonas* following diverse lifestyles. *Xylella* is a lethal plant pathogen with highly reduced genome, atypical GC content and is taxonomically related to these three genera. Deep phylo-taxono-genomics reveals that *Xylella* is a variant *Xanthomonas* lineage that is sandwiched between *Xanthomonas* species. Comparative studies suggest the role of unique pigment and exopolysaccharide gene clusters in the emergence of *Xanthomonas* and *Xylella* clades. Pan genome analysis identified set of unique genes associated with sub-lineages representing plant associated *Xanthomonas* clade and nosocomial origin *Stenotrophomonas*. Overall, our study reveals importance to reconcile classical phenotypic data and genomic findings in reconstituting taxonomic status of these four genera.

Significance Statement

Xylella fastidiosa is a devastating pathogen of perennial dicots such as grapes, citrus, coffee, and olives. The pathogen is transmitted by an insect vector to its specific host wherein the infection leads to complete wilting of the plants. The genome of *X. fastidiosa* is extremely reduced both in terms of size (2Mb) and GC content (50%) when compared with its relatives

such as *Xanthomonas, Stenotrophomonas, and Pseudoxanthomonas* that have higher GC content (65%) and larger genomes (5Mb). In this study, using systematic and in-depth genome-based taxonomic and phylogenetic criteria along with comparative studies, we assert the need of unification of *Xanthomonas* with its misclassified relatives (*Xylella*, *Stenotrophomonas* and *Pseudoxanthomonas*). Interestingly, *Xylella* revealed itself as a minor lineage embedded within two major *Xanthomonas* lineages comprising member species of different hosts.

Introduction

Family *Lysobacteraceae* (*Xanthomonadaceae*) (CHRISTENSEN & Cook, 1978; S. Naushad et al., 2015) harbors bacterium of diverse ecological niches. Within this family, *Xanthomonas*, *Stenotrophomonas*, *Xylella* and *Pseudoxanthomonas* are closely related genera which forms a phylogroup (referred to as XSXP phylogroup in this study) (Kumar, Bansal, Patil, & Patil, 2019). *Xanthomonas*, *Xylella* and *Pseudoxanthomonas* are characterized by a yellow pigment xanthomonadin and exopolysaccharide xanthan gum production (Biswas, Chakraborty, Sarkar, & Naidu, 2017; da Silva, Vettore, Kemper, Leite, & Arruda, 2001; He, Cao, & Poplawsky, 2020; Katzen et al., 1998; Lu et al., 2008; Rajagopal, Sundari, Balasubramanian, & Sonti, 1997). XSXP phylogroup have a long history of taxonomic reshuffling based on phenotypic, morphological characteristics and genotypic methods. These genotypic methods were based on single gene such as 16S rRNA, *rpoB* or *gyrB* gene or multiple housekeeping genes (Parkinson, Cowie, Heeney, & Stead, 2009; Saddler & Bradbury, 2005; Yarza et al., 2010; Yilmaz et al., 2014). Taxonomic and phylogenetic relationship amongst XSXP based on classical methods have been always debatable. For instance, basionym of *Xanthomonas* and *Stenotrophomonas* were *Bacillus campestris* by Pammel in 1895 (Pammel, 2017) and *Pseudomonas maltophilia* respectively (Hugh & Ryschenkow, 1961). Further, *Bacillus campestris* including few groups of the plant pathogens were later classified as *Xanthomonas campestris* (Dowson, 1939). Similarly, *Pseudomonas maltophilia* was transferred to genus *Xanthomonas* as *X. maltophilia* (Swings, De Vos, den MOOTER, & De Ley, 1983) and later it was designated as a distinct and new genus *Stenotrophomonas* (Palleroni & Bradbury, 1993). Similarly, *Xylella* (Wells et al., 1987) and *Pseudoxanthomonas* (Finkmann, Altendorf, Stackebrandt, & Lipski, 2000) were defined as species related to *Xanthomonas*. However, these taxonomic assignments remain inconclusive as these were based on few evidences of classical taxonomy such as 16S rRNA identity or biochemical properties etc. Current standing in nomenclature of *Lysobacteraceae* (https://lpsn.dsmz.de/family/lysobacteraceae) classifies XSXP phylogroup as four distinct

genera (i.e., *Xanthomonas, Pseudoxanthomonas, Stenotrophomonas* and *Xylella*). This is based on conserved sequence indels (CSI) and phylogeny of only 28 proteins (H. S. Naushad & Gupta, 2013; S. Naushad et al., 2015). However, CSI phylogeny also suggests intermingling of some species of these four genera (H. S. Naushad & Gupta, 2013; S. Naushad et al., 2015). Further, genome taxonomy database (GTDB) also supports inclusion of several genera in the XSXP phylogroup. GTDB classification is based on 120 conserved genes and relative evolutionary divergence values. However, incongruence to the GTDB proposed taxonomy is reported in some previous focused studies (Liao, Lin, Li, Qu, & Tian, 2020; Zheng et al., 2020). GTDB proposed classification for XSXP phylogroup is not yet reported in literature and hence, its evolutionary and taxonomic status is worth discussing.

Taxonomy and phylogeny go hand in hand. Amongst XSXP, existence of two major phylogroups within genus *Xanthomonas* (group 1 and group 2) appears as early as in 1960s (Colwell & Liston, 1961). Later on, single gene phylogeny of 16S rRNA and *gyrB* (Hauben, Vauterin, Swings, & Moore, 1997; Parkinson et al., 2007; Pieretti et al., 2009; Studholme et al., 2011) defined the group 1 as early branching. Group 1 is highly diverse which constitutes species like *X. albilineans* with reduced genome (Pieretti et al., 2009), *X. sontii* following non-pathogenic lifestyle (Bansal, Kaur, et al., 2019; Bansal, Kumar, & Patil, 2020; Bansal, Midha, et al., 2019a) in addition to several pathogens like *X. sacchari, X. translucens*, etc. Whereas, group 2 is the largest and constitutes well-described pathogens such as *X. oryzae, X. citri, X. campestris* etc. (Hauben et al., 1997; Parkinson et al., 2007).

However, phylogeny of two devastating phytopathogenic genera *Xanthomonas* and *Xylella* on the basis of single or multi-locus housekeeping genes have been debatable over the years (Pieretti et al., 2009; Ryan et al., 2011; Studholme et al., 2011). These studies did not resolve their relationship and either places *Xylella* within group 1 and group 2 (Pieretti et al., 2009) of *Xanthomonas* or as a distinct monophyletic lineage (Ryan et al., 2011). Furthermore, 16S rRNA based phylogeny have also placed *Xanthomonas* and *Stenotrophomonas* as coherent group excluding *Xylella* (Pieretti et al., 2009).

With recent advancements, genomics is at the center of revolution in bacterial taxonomy and phylogeny. Whole genome comparisons of *Xanthomonas* and *Xylella* have revealed high degree of identity and co-linearity of their chromosome backbone (Monteiro-Vitorello et al., 2005; Van Sluys et al., 2002). Even though, their genomes have diverged by potential indels mediated by mobile genetic elements, *Xylella* shares 74% of the genes with *Xanthomonas* (Moreira et al., 2004) (Monteiro-Vitorello et al., 2005). *Xylella* genome size is roughly two-

third of the *Xanthomonas* genome with minimal complement of genes required for its survival within host (Lu et al., 2008). Like *Xylella*, *Xanthomonas albilineans* is also a xylem-limited plant pathogen with reduced genome and characterized by the absence of Hrp-T3SS (hypersensitive response and pathogenicity–type III secretion system (Pieretti et al., 2009). However, genome reduction events in genus *Xanthomonas* are not limited to group 2 (*Xanthomonas albilineans*) but also reported in group 1 i.e., *Xanthomonas vasicola* (Rodriguez-R et al., 2012). All these preliminary genome level investigations based on limited strains provide certain clues of *Xanthomonas* and *Xylella* relatedness. Yet small genome size, low GC content and dual lifestyle of *Xylella* has compelled taxonomists to consider it as a different genus (S. Chatterjee, R. P. Almeida, & S. Lindow, 2008a). However, in microbial world genome erosion cannot warranty for a new genus rather it leads to a highly specialized pathogen. For instance, *Mycobacterium leprae* having almost 50% of the genome reduced as compared to *M. tuberculosis* (Gómez-Valero, Rocha, Latorre, & Silva, 2007).

Phylogenetic investigation which led to current standing in nomenclature was based on just 28 conserved proteins and that too of less number of species of XSXP phylogroup. Moreover, intermingled phylogeny of *Xanthomonas* and *Xylella* in the CSI phylogeny was also overlooked in the current standing in nomenclature (https://lpsn.dsmz.de/family/lysobacteraceae) (H. S. Naushad & Gupta, 2013; S. Naushad et al., 2015). Hence, this does not provide true evidence for considering XSXP phylogroup as four distinct genera. The increasing evidences of relatedness amongst the XSXP phylogroups beyond the genus level is gradually strengthening. However, to get robust taxonomy on the basis of phylogenomic framework a comprehensive genome level investigation considering all representative species of XSXP is required. Present deep whole genome based phylogenetic and evolutionary investigation reveals that all the four genera indeed belong to the same genus *Xanthomonas*. Such a deep phylo-taxono-genomic or DEEPT genomic study of XSXP have major implications in understanding the co-evolution of microbes with their host plants.

Results

Genomic features of XSXP phylogroup

Genomic features of type strains of the XSXP phylogroup are summarized in table 1. Genome size of *Xanthomonas* is around 5 Mb and around 3-5 Mb for *Pseudoxanthomonas* and *Stenotrophomonas*, whereas, genome size of *Xylella* is around 2.5 Mb. This reduction in

genome size is also reflected in number of coding sequences that is in the range of 3000 to 4000 for *Xanthomonas, Stenotrophomonas* and *Pseudoxanthomonas,* but around 2100 for *Xylella.* Furthermore, average GC content of the genome for all the members of *Xanthomonas, Stenotrophomonas* and *Pseudoxanthomonas* is ~65% except for *Xylella* that has a GC content of around 51%. In spite of dramatic genome reduction *Xylella* is displaying comparable number of tRNAs i.e., 48.

Phylogenomics of XSXP phylogroup

To investigate the relationship amongst the member species of XSXP phylogroup, we constructed and compared genome-based phylogeny by three different approaches. Based on an earlier study of family with basonym *Xanthomonadaceae* (*Lysobacteraceae*) and order with basonym *Xanthomonadales* (*Lysobacterales)*, we used *Luteimonas mephitis* DSM12574^T as an outgroup of XSXP (Kumar et al., 2019). Phylogeny constructed (including XSXP phylogroup and *Luteimonas mephitis* DSM12574^T as outgroup) on the basis of large set of genes core to the bacterial world (>400 genes) (figure 1a) and set of genes core to XSXP phylogroup by roary (382 genes across 99-100% isolates) (Page et al., 2015) and PIRATE (1149 genes across 95-100% isolates) (Bayliss, Thorpe, Coyle, Sheppard, & Feil, 2019) (figure 1b, supplementary figure 1) correlated with each other (pl. see methods). Core genome level phylogeny revealed that among XSXP phylogroup, *Pseudoxanthomonas* members were more diverse and ancestral to other three genera. Interestingly, other three genera formed two mega species groups (MSG) i.e. plant pathogens *Xanthomonas* and *Xylella* comprised one MSG referred as XX-MSG and *Stenotrophomonas* formed another MSG referred as S-MSG. XX-MSG consist of three clades. Clade I comprises of at least 27 species that are primarily pathogens of dicot plants which was earlier considered as group 2 (Hauben et al., 1997; Parkinson et al., 2007), clade III comprises of at least 6 species that are primarily pathogen of monocot plants including *X. albilineans* with reduced genome and *X. sontii* with non-pathogenic lifestyle earlier considered as group 1. Clade II comprised both the species of *Xylella* sandwiched in between clade I and clade III of genus *Xanthomonas.* Members of clade I are pathogens of dicots, except 5 species which have monocot hosts i.e. *X. vasicola, X. oryzae, X. axonopodis, X. bromi* and *X. maliensis*. On the other hand, members of clade III are pathogens or associated with monocots, except *X. theicola* infecting dicots. Interestingly, *Stenotrophomonas panacihumi*, an environmental species, reflected as singleton phylogenetic outlier of clade I and clade II. In case of S-MSG, we found clade IV comprising of *Stenotrophomonas maltophilia* complex (Smc) of clinical origin.

Taxonogenomics of XSXP phylogroup

In order to delineate members of XSXP at genera and species level, we further carried out taxonogenomic analysis. Apart from the phylogenetic trees based on large sets of genomic markers and core genome, new criteria are also becoming available for delineating members at the genus level. For genus demarcation 73.98% average nucleotide identity (ANI) and 0.33 alignment fraction (AF) are cut-offs set which are obtained by large scale genus comparisons (Auch, Jan, Klenk, & Göker, 2010; Barco et al., 2020; Richter & Rosselló-Móra, 2009). Further, criteria such as average amino acid identity (AAI) and percentage of conserved protein (POCP) have been proposed for genus delineation with 60-80% and 50% cut-offs, respectively (Konstantinidis & Tiedje, 2005; Qin et al., 2014). POCP is affected by genome size and hence cannot be applied for *Xylella* that has undergone extreme genome reduction (Hayashi Sant'Anna et al., 2019; Qin et al., 2014). However, core genome-based phylogenetic trees and AAI are not affected by genome reduction and can be employed to establish the identity and relationship of genera irrespective of genome size and/or GC content (Hayashi Sant'Anna et al., 2019; Indu, Ch, & Ch V, 2019; Konstantinidis & Tiedje, 2005). Interestingly, according to the genus demarcation values, minimum AF and ANI values for the XSXP phylogroup excluding *Xylella* were 0.35 and 76.54% respectively depicting *Xanthomonas, Stenotrophomonas* and *Pseudoxanthomonas* belong to same genus (supplementary table 1). Further, according to AAI cut-off value also all the strains of XSXP belong to the same genus i.e., *Xanthomonas* (figure 2a) and hence need to be merged into a single genus as *Xanthomonas.* Similarly, according to POCP cut-off of 50% all members of *Xanthomonas, Stenotrophomonas* and *Pseudoxanthomonas* belong to a single genus (figure 2b). Furthermore, within XX-MSG, 27 species formed clade I while other 6 species formed clade III. Further, ANI values clearly depicted that all the members of clade I and clade III belong to different species except for *X. alfalfa, X. perforans, X. euvesicatoria* and *X. gardneri, X. cynarae,* which represent heterotypic synonyms (supplementary figure 2). Clade II composed of *Xylella* genus. In case of S-MSG, the *Stenotrophomonas maltophilia* complex (Smc) consist of 6 species with two of them being clinical origin i.e. *S. maltophilia* and *S. africana*.

Pan genome analysis and genome dynamics in XSXP

We performed pan genome analysis using roary v3.12.0 (Page et al., 2015) with 50% or greater amino acid identity revealed 47,624 gene families and 397 core genes for the XSXP phylogroup. Further, 48, 53, 205 and 175 genes were unique to clade I, III, II and IV

respectively (supplementary table 2, 3, 4 and 5). Interestingly, clades II and IV constituting *Xylella* and Smc respectively are having higher number of unique genes. Inspection of clade IV unique genes revealed a type II secretion system, peroxidase, peptidases, efflux pumps and transporters like antibiotics/antimicrobials, fluoride ions, solvent, TonB dependent receptors, transcriptional regulators etc. (supplementary table 4). Further, unique genes of clade II belong to adhesin like type IV pili formation apart from glycosyltransferases, methyltransferases etc. (supplementary table 3). Overall COG classification revealed that hypothetical proteins are dominant class in all the clades suggesting unknown functions playing role in their success. Interestingly, in both clade I and III, COG class related to "signal transduction mechanisms" is second major class after hypothetical proteins, clade II is having more of "cell wall/membrane/envelope biogenesis", "secondary metabolites biosynthesis, transport and catabolism" and "coenzyme transport and metabolism". Whereas, clade IV was having more of "transcription", "signal transduction mechanisms", "intracellular trafficking, secretion and vesicular transport", "cell wall/membrane/envelope biogenesis" and "inorganic ion transport and metabolism" (figure 3).

Variable patterns of recombination within XSXP

To characterize genome-wide mosaicism in XSXP phylogroup we ran fastGEAR (Mostowy et al., 2017) on individual sequence alignments of the core and accessory genes or the pan genes. We found that 5,434 genes out of 47,162 pan genes had experienced recombination representing 11.5% of the pangenome (supplementary table 6). Here, 53,125 were the total recombining events detected out of which only 30% were detected from XX-MSG and remaining 70% were detected in genera *Pseudoxanthomonas* and *Stenotrophomonas* (figure 4). Amongst XX-MSG, least recombination events (72 out of 53125) were detected in *Xylella.* We observed heterogeneity in recombination sizes, majority of the recombination (~82%) were of less than 100bp recombining fragments while maximum recombining fragment was of 4469bp (supplementary table 6, 8). Further, 14 highway pairs were detected in the XSXP phylogroup, majority related to *Pseudoxanthomonas* and *Stenotrophomonas* (figure 5, supplementary table 7). Here, highways likely to represent specific lineages that function as hubs of gene flow. Here, highest recombination events were detected in *Pseudoxanthomonas* while least in *Xylella* strains.

Xanthomonadin pigment and xanthan exopolysaccharide gene cluster

Distinct yellow xanthomonadin pigment, responsible for yellow-colored colonies, encoded by *pig* gene cluster, and the thick mucus exopolysaccharide or xanthan encoded by *gum* gene cluster are characteristic features of canonical plant associated *Xanthomonas* species (He et al., 2020; Katzen et al., 1998; Rajagopal et al., 1997). Since, we are expanding breadth of *Xanthomonas* genus on the basis of phylo-taxono-genomics parameters, we scanned for the presence of these clusters in the genomes of XSXP constituent genera and species members (figure 6). Among the XX-MSG, the pig and gum gene clusters are present in all the members of *Xanthomonas* and *Xylella*, except for *X. theicola* and *X. albilineans* which are not having gum gene cluster. While both the clusters are highly conserved in sequence and distribution in clade I and II members of clade II show incomplete and possibly degenerated clusters, with *gumN, gumM, gumI, gumG* and *gmuF* absent from the *gum* cluster and orfs 8, 11, 13 and 14 absent from the pigment cluster.

In members of other two genera understudy (*Stenotrophomonas* and *Pseudoxanthomonas*), these clusters are either absent or incomplete in majority of the members. Specifically, while all members of genus *Stenotrophomonas* lacks *gum* gene cluster, and 5 out of 20 species of *Pseudoxanthomonas* i.e., *P. jiangsuensis, P. composti, P. spadix, P. wuyuanensis* and *P. sacheonensis* harbour *gum* gene cluster. Whereas, xanthomonadin cluster is widely present in other two genera i.e., 7 out of 19 *Stenotrophomonas* and 13 out of 20 *Pseudoxanthomonas* species (figure 6). Remarkably, none of the clade IV of some clinical isolates have both the clusters.

Discussion

The ability to sequence the genomes of bacterial strains in a cost-effective and high-throughput manner is transforming the way we understand genetics, phylogeny, and taxonomy. Inferring the phylogeny based on limited and highly-conserved sequence information such as 16S rRNA and housekeeping genes can be misleading not only at the species level but also at higher taxonomic levels (Sangal et al., 2016). Genomic investigation is the robust way to establish an organism's identity, biology, and ecology in the proper context. Genome-based comprehensive taxonomic studies at various levels, such as order *Methylococcales*, order *Bacillales*, genus *Borrelia, and* genus *Lactobacillus*, have resolved the relationships and provided a robust basis for reclassification (De Maayer, Aliyu, & Cowan, 2019; Margos et al., 2019; Orata, Meier-Kolthoff, Sauvageau, & Stein, 2018; Salvetti, Harris, Felis, & O'Toole, 2018; Zheng et al.,

2020). Whole genome studies using representative reference strains and type species of genera allowed us to resolve misclassifications at the level of families in the order *Lysobacterales* (*Xanthomonadales)* (Kumar et al., 2019). Whole genome studies also allowed the resolution of misclassifications at the level of species and clones (Bansal, Kumar, & Patil, 2021; Bansal, Midha, Kumar, & Patil, 2017; Kumar et al., 2019). In the case of genus *Stenotrophomonas*, we reported a large and hidden *Stenotrophomonas maltophilia* complex (Smc) consisting of at least nine species (Prashant P Patil, Kumar, Midha, Gautam, & Patil, 2018). Smc is a dynamically evolving group of human opportunistic pathogens with extreme drug resistance (Gautam et al., 2021; Kumar et al., 2020). However, deeper inter-genera and intra-genera genome-based investigations at the phylogenetic and taxonomic levels are lacking for the XSXP phylogroup. Considering the importance of the members of the XSXP phylogroup for plant and human health apart from their biotechnological potential, we performed an in-depth investigation by incorporating all the representatives of four closely-related genera through modern genome-based criteria.

Such a comprehensive phylo-taxono-genomic study revealed the role of evolutionary and ecological diversity in the formation of clades that actually belong to the genus *Xanthomonas* but have been historically classified into distinct genera such as *Xylella*, *Stenotrophomonas,* and *Pseudoxanthomonas*. The basal, multiple, and diverse groups comprising species of the genus *Pseudoxanthomonas* suggest it to be ancestral to the other three genera. Hence, it is not surprising that this large genus is named as *Pseudoxanthomonas* (Finkmann et al., 2000). At the same time, the presence of major lineage comprising the member species of *Stenotrophomonas* and another consisting of *Xanthomonas-Xylella* species suggests further specialization based on habitats or ecology. This phenomenon is quite obvious in the canonical plant-associated and pathogenic *Xanthomonas-Xylella* species that form a distinct mega-group unlike the mega-species group(s) represented by member species of *Stenotrophomonas* and *Pseudoxanthomonas* whose members are highly versatile. Besides, this finding indicates the role of lifestyle in the diversification and formation of clades that were previously reflected as separate taxonomic units at the level of genera when using only limited genotypic and phenotypic data. More importantly, we were also able to establish the robust phylogenetic relationship among the constituent member genera and species in XSXP using multiple phylogenetic approaches. In our case, this was possible because of the phylogenetic foundation provided by the species of *Pseudoxanthomonas* for the other three genera, asserting the

significance of the genomic resource of the former. This proves the power of deep phylogenetic studies covering all member species and closely related genera.

Establishing the position of *Xylella*, which has a fastidious nature and an extremely lowered GC content along with highly reduced genome size, is of critical importance. While it is obvious that at the taxonogenomic level all the four genera are not distinct but belong to one genus, the element of surprise is that the genus *Xylella* is more related to *Xanthomonas* than to *Stenotrophomonas*. This variant lineage was confirmed by the sandwiched phylogenomic position of *Xylella* within the XX-MSG. The event that led to a sudden decrease in the GC content by >10% and genome size by ~50% is difficult to discern, and further studies in this regard are required. Our DEEPT studies imply that such an event could have happened during the diversification of *Xanthomonas* into pathogen of dicots from monocots or during diversification of *Xanthomonas* into a pathogen of monocots and dicots.

Conservation of the xanthomonadin pigment and *gum* gene clusters in clade I and clade III suggest to the essentiality of the pigment and the exopolysaccharide in the obligate plant-associated lifestyle. Since plants are directly and continuously exposed to light, it is important for a successful phytopathogen to have a unique pigment such as xanthomonadin (Rajagopal et al., 1997). It has been shown that xanthomonadin provides protection against photodamage in *X. oryzae* pv. oryzae that infects rice (Rajagopal et al., 1997). As plants are also known to produce large amounts of antimicrobial compounds and secondary metabolites (Ramírez-Gómez, Jiménez-García, Campos, & Campos, 2019), there is a need for a unique and thick polysaccharide coat. In fact, highly mucoid and distinct yellow-coloured colonies support our hypothesis that these clusters are evolved in the canonical *Xanthomonas* species from XX-MSG. Both these clusters were found to be more diversified in clade III than in clade I, which can be linked to the majorly dicot-associated lifestyle of the latter as compared to the majorly monocot associated lifestyle of the former. The leaves of monocots and dicots are different, which may affect the penetration of light and its effect on bacteria. Nonetheless, this hypothesis needs to be proven by further studies. Both the clusters were degenerated or incomplete in case of clade II, which can be associated with the dual lifestyle of *Xylella* that occurs within the foregut of the insect and within the plant where it is directly injected into the xylem (S. Chatterjee, R. P. P. Almeida, & S. Lindow, 2008b). Since *Xylella* is never exposed to light and plant defence response unlike the canonical *Xanthomonas* (i.e. clade I and III), it is not

surprising that its clusters are not under natural selection. Genetic studies have revealed that the *pig* cluster is not required for virulence in plants (Rajagopal et al., 1997). Hence, the conservation of this cluster in clade I and III in the absence of its role in virulence suggests that the epiphytic mode is also crucial in plant adaptation (Pandey & Sonti, 2010).

The distributed presence of *pig* and *gum* clusters in a few members of XSXP outside the XX-MSG suggests that these clusters are ancestral and were present as incomplete clusters in the population even before the emergence of *Xanthomonas* clades. Eventually, they must have undergone diversification and selection when a member of XSXP came in contact with the plants. The ancestor of XX-MSG probably possessed both the clusters, thereby conferring an advantage for the plant-associated lifestyle. However, the presence of a primitive, ancestral, and incomplete pigment and *gum* gene clusters points that these clusters are not of foreign origin but were inherent to the ancestral XSXP population.

Even though *Xylella* seems to have undergone drastic genome reduction, the acquisition of a large number of unique genes with particular functions has played a key role in its emergence as a successful phytopathogen with insect/plant dual lifestyle. In both these variant lineages i.e. clade II and IV, the gaining of functions related to regulation suggest that apart from gene gain, the regulation of core or novel genes has also been responsible for their success. The commonality of the unique genes with regard to functions such as “intracellular trafficking, secretion, and vesicular transport”, “defence mechanism”, “replication, recombination, and repair” and “cell cycle control, cell division, and chromosome partitioning” reiterate their opportunistic, variant origin and parallels in evolution for dual/opportunistic lifestyle. Even though both exhibit dual lifestyle, unique gene analysis allowed us to pinpoint the functions associated with their success, such as adhesion in *Xylella* and efflux proteins/peroxidise (for multidrug resistance, adaptation to hospital setting, etc.) along with a novel type II secretion system in Smc. In an opportunistic human pathogen, a novel type II secretion system may compensate for the absence of type III secretion system in Smc (Crossman et al., 2008).

Even within the *Stenotrophomonas* group or S-MSG, the formation of a species complex corresponding to the MDR nosocomial pathogen points to the ongoing evolution and diversification as witnessed in the case of *Xylella* within XX-MSG. This finding is valuable in furthering our understanding of this emerging opportunistic human pathogen. Unique genes in clade I and III may be related to their adaptation to dicots and monocots, respectively. While many unique genes encode hypothetical proteins suggesting the importance of further

functional genetic studies in clade I, II, III and IV, other major classes also provide clues regarding their evolution. One such hint is the importance of signal transduction in clade I and III, which can then serve as key targets for both basic and applied studies of dicot and monocot pathogens. Further systematic gene content analysis by excluding the variants within this group will also allow us to obtain insights into those genes that are important for the adaptation of *Xanthomonas* to monocots and dicots. Overall, our study reiterates the power and potential of systematic and large scale or deep taxonogenomics.

Reconciliation, reconstitution and emended description of the genus *Xanthomonas*

Taxonomy in the pre-genomic era was dominated by classical polyphasic studies like phenotypic, microscopic, biochemical, fatty acids, quinines, etc. Any slight variation in the phenotypic data had led to assignment of new genus amongst *Xanthomonas* like species such as *Pseudomonas* (Hugh & Ryschenkow, 1961; Swings et al., 1983), *Stenotrophomonas (Palleroni & Bradbury, 1993), Pseudoxanthomonas* (Finkmann et al., 2000) and *Xylella* (Wells et al., 1987). *Pseudoxanthomonas* shares phenotypic traits with *Xanthomonas* like species such as colony color, cell shape, Gram staining, presence of branched chain fatty acid pattern and ubiquinone with eight isoprenoid units. Yet, *Xanthomonas* like strains with the ability to reduce nitrite but not nitrate to N_2O and by the lack of C13 : 0 iso 3-OH fatty acid were designated as genus *Pseudoxanthomonas* (Finkmann et al., 2000; Yang, Vauterin, Vancanneyt, Swings, & Kersters, 1993). Similarly, xylem-limited *Xanthomonas* like fastidious strains with GC content 51 to 53 mol % were designated as separate genus *Xylella* (Wells et al., 1987). Presence of fimbrae, multiple flagella and inability to produce xanthan gum and xanthomonadins distinguished another new genus *Stenotrophomonas* amongst the *Xanthomonas* like species (Palleroni & Bradbury, 1993).

Even then, these genera have characteristic common presence of branched chain fatty acid pattern and an ubiquinone with eight isoprenoid units (Q-8) (Finkmann et al., 2000; Swings et al., 1983). Hence, classification of *Xanthomonas* like species as distinct genera has been controversial which led to several reclassifications (Lee et al., 2008; Palleroni & Bradbury, 1993). Also plant or human associated microbes have influenced their assignment into separate genera. Advent of genomics have highlighted role of horizontal gene transfer in diverse lifestyles of these microbes. Acquisition of genes would have provided much needed functions

in adaption to humans (in case of *S. maltophilia* complex) (Prashant P. Patil et al., 2021) and insect host (in case of *Xylella*) (Chatterjee et al., 2008b).

Hence, there is a need to reconcile the published phenotypic data in the light of genomic data and relationship of all the four genera. The current standing in nomenclature of family *Lysobacteraceae* is based on CSI phylogeny (using only 28 conserved proteins) which describes these *Xanthomonas* like species into distinct genera namely *Xanthomonas, Xylella, Stenotrophomonas* and *Pseudoxanthomonas* (H. S. Naushad & Gupta, 2013; S. Naushad et al., 2015). Although CSI phylogeny itself suggests taxonomic position of *Xylella* in between *Xanthomonas* group 1 and group 2 which is not a true evidence for considering *Xanthomonas* and *Xylella* as distinct genera (S. Naushad et al., 2015). Hence, there is a need to consider them as synonyms of genus *Xanthomonas*.

In the present study we provide whole genome based deep-phylo-taxonogenomic evidences which clearly depicts the existence of unary genus for all *Xanthomonas* like species (*Xanthomonas, Xylella, Stenotrophomonas* and *Pseudoxanthomonas*). Here, we have implemented three different robust approaches to obtain core-genome phylogeny of XSXP phylogroup i.e. PhyloPhlAn (>400 genes), roary (382 across 99-100% isolates) and PIRATE (1149 genes across 95-100% isolates). Core genome phylogeny obtained from these methods reveals sandwiched positioning of *Xylella* between two groups of *Xanthomonas.* While, *Pseudoxanthomonas* was ancestral among XSXP phylogroup. Further, genus delineation using ANI, AF, AAI, POCP clearly depicted *Xanthomoans, Stenotrophomonas* and *Pseudoxanthomonas* as single genus. However, *Xylella* cannot be assessed using genome similarity assessment due to its drastic reduction in genome size (leading to low number of CDS and GC content). In light of deep phylo-taxono-genomics findings along with published polyphasic data XSXP phylogroup warrants reunification of all its *Xanthomonas* like species.

Materials and Methods

Genome procurement from public repository

A total of 76 strains were used in the present study. Genome sequences of 33, 20, 20 and 3 were type strains of *Xanthomonas*, *Stenotrophomonas* and *Xylella*, were available from the NCBI database (https://www.ncbi.nlm.nih.gov/). All the genomes were then accessed using

checkM v1.1.0 (Parks, Imelfort, Skennerton, Hugenholtz, & Tyson, 2015) with the cutoff of less than 3 percent contamination and completeness (table 1).

Whole genome-based phylogeny

Phylogenomic analysis based on more than 400 putative conservative genes was carried out using PhyloPhlAn v3.0 (Asnicar et al., 2020). Here, USEARCH v5.2.32 (Edgar, 2010), MUSCLE v3.8.3 (Edgar, 2004) and FastTree v2.1 (Price, Dehal, & Arkin, 2010) were utilized for orthologs searching, multiple sequence alignment and phylogenetic construction respectively. All strains of XSXP phylogroup and *Luteimonas mephitis* DSM12574^T (used as an outgroup) were used for construction of the phylogeny.

In order to obtain a more robust whole genome phylogeny, core genome based phylogeny was constructed using MAFFT v7.31 (Nakamura, Yamada, Tomii, & Katoh, 2018) (https://mafft.cbrc.jp/alignment/software/) and the FastTree v2.1 (Price et al., 2010) which was integrated within the roary v 3.12.0 (Page et al., 2015) with identity cut-off of 60%. Implementation of the tools can be found in detail under the section of pan genome analysis.

As here, XSXP phylogroup consist of divergent genomes and to further confirm the core genomes deduced by roary and their phylogeny, we have used PIRATE (Bayliss et al., 2019). This is better known to identify orthologue groups of diverged gene sequences by using amino acid identity threshold of 50%, 60%, 70%, 80%, 90% and 95%.

Taxonogenomic analysis

Genome relatedness was assessed using the average amino acid identity (AAI), percentage of conserved protein (POCP) and average nucleotide identity using (ANI). AAI was calculated using CompareM v0.0.23 (https://github.com/dparks1134/CompareM), which uses the mean amino acid identity of orthologous genes between a given pair of genomes. POCP is another method to evaluate the genome relatedness at genus level, which is based on amino acid conservation. POCP is calculated with the blast search using the default settings (Qin et al., 2014) (https://figshare.com/articles/POCP-matrix_calculation_for_a_number_of_genomes/5602957). Further, FastANI v1.3 (Jain, Rodriguez-R, Phillippy, Konstantinidis, & Aluru, 2018) an alignment-free sequence mapping method with default settings was used to calculate ANI values with default settings.

Pan genome analysis

Pan genome analysis of the strains was performed using Roary v3.12.0 (Page et al., 2015). Here, gff files generated by Prokka v1.13.3 (Seemann, 2014) were used as input for Roary pan genome analysis. Since, we are analyzing genomes from different species, the identity cut-off used was 60%. Functional annotation of the core genes identified was performed using eggNOG v4.5.1 (Huerta-Cepas et al., 2015).

Recombination analysis

To identify recombination, we used fastGEAR (Mostowy et al., 2017)with default parameters on individual genes of the pan genome identified by Roary. HERO uses output of fastGEAR to identify donors and recipients in the recombination events. It also mapped recombination events to the sequence clusters identified by core genome-based phylogeny (supplementary table 6) to elucidate potential drivers of biases in recombination partners.

The custom script has been provided with HERO on its GitHub page (https://github.com/therealcooperpark/hero) as "sidekick.py" for reproducibility and convenience when using HERO in similar workflows.

Cluster analysis

Protein sequences of the gene clusters (Bansal, Midha, et al., 2019b) were used as query and tBLASTn searches were performed on the XSXP phylogroup genomes. Here, tBLASTn searches were performed using standalone BLAST+ v2.9.0 (Camacho et al., 2009) and cut-offs used for identity and coverage were 50% and 50% respectively. Heatmap for the blast searches were generated using GENE-E v3.03215 (https://software.broadinstitute.org/GENE-E/).

Author Contributions

KB and SK carried out all the bioinformatics analysis. KB, SK and PBP drafted the manuscript with inputs from AK and AS. PBP has conceived the study and participated in the design. All the authors have read and approved the manuscripts.

Acknowledgement

This work was supported by a project entitled "Megagenomic and Metagenomic insights into adaptation and evolution of fruit Microbiome" GAP0187 of CSIR to P.B.P.

References

Asnicar, F., Thomas, A. M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., . . . May, U. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nature communications, 11*(1), 1-10.

Auch, A. F., Jan, M., Klenk, H.-P., & Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in Genomic Sciences, 2*(1), 117.

Bansal, K., Kaur, A., Midha, S., Kumar, S., Korpole, S., & Patil, P. B. (2019). Xanthomonas sontii sp. nov., a non-pathogenic bacterium isolated from healthy basmati rice (Oryza sativa) seeds from India. *bioRxiv, bioRxiv*, 738047.

Bansal, K., Kumar, S., & Patil, P. B. (2020). Phylogenomic Insights into Diversity and Evolution of Nonpathogenic Xanthomonas Strains Associated with Citrus. *Msphere, 5*(2).

Bansal, K., Kumar, S., & Patil, P. B. (2021). Taxonomic repositioning of twelve Xanthomonas campestris, seven Xanthomonas axonopodis and one Pseudomonas cissicola reference pathovars to Xanthomonas citri. *bioRxiv*.

Bansal, K., Midha, S., Kumar, S., Kaur, A., Sonti, R. V., & Patil, P. B. (2019a). Ecological and evolutionary insights into pathogenic and non-pathogenic rice associated Xanthomonas. 453373.

Bansal, K., Midha, S., Kumar, S., Kaur, A., Sonti, R. V., & Patil, P. B. (2019b). Ecological and evolutionary insights into pathogenic and non-pathogenic rice associated Xanthomonas. *bioRxiv*, 453373.

Bansal, K., Midha, S., Kumar, S., & Patil, P. B. (2017). Ecological and evolutionary insights into Xanthomonas citri pathovar diversity. *Applied and environmental microbiology, 83*(9), e02993-02916.

Barco, R., Garrity, G., Scott, J., Amend, J., Nealson, K., & Emerson, D. (2020). A genus definition for bacteria and archaea based on a standard genome relatedness index. *MBio, 11*(1), e02475-02419.

Bayliss, S. C., Thorpe, H. A., Coyle, N. M., Sheppard, S. K., & Feil, E. J. (2019). PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. *GigaScience, 8*(10), giz119.

Biswas, B., Chakraborty, A., Sarkar, B., & Naidu, R. (2017). Structural changes in smectite due to interaction with a biosurfactant-producing bacterium Pseudoxanthomonas kaohsiungensis. *Applied Clay Science, 136*, 51-57.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC bioinformatics, 10*(1), 421.

Chatterjee, S., Almeida, R. P., & Lindow, S. (2008a). Living in two worlds: the plant and insect lifestyles of Xylella fastidiosa. *Annual review of phytopathology, 46*.

Chatterjee, S., Almeida, R. P. P., & Lindow, S. (2008b). Living in two worlds: the plant and insect lifestyles of Xylella fastidiosa. *Annu. Rev. Phytopathol., 46*, 243-271.

CHRISTENSEN, P., & Cook, F. (1978). Lysobacter, a new genus of nonfruiting, gliding bacteria with a high base ratio. *International Journal of Systematic Evolutionary Microbiology, 28*(3), 367-393.

Colwell, R., & Liston, J. (1961). Taxonomic analysis with the electronic computer of some Xanthomonas and Pseudomonas species. *Journal of bacteriology, 82*(6), 913.

Crossman, L. C., Gould, V. C., Dow, J. M., Vernikos, G. S., Okazaki, A., Sebaihia, M., . . . Peters, N. (2008). The complete genome, comparative and functional analysis of Stenotrophomonas maltophilia reveals an organism heavily shielded by drug resistance determinants. *Genome biology, 9*(4), R74 %@ 1474-1760X.

da Silva, F. R., Vettore, A. L., Kemper, E. L., Leite, A., & Arruda, P. (2001). Fastidian gum: the Xylella fastidiosa exopolysaccharide possibly involved in bacterial pathogenicity. *FEMS Microbiology Letters, 203*(2), 165-171.

De Maayer, P., Aliyu, H., & Cowan, D. A. (2019). Reorganising the order Bacillales through phylogenomics. *Systematic and applied microbiology, 42*(2), 178-189.

Dowson, W. (1939). On the systematic position and generic names of the Gram-negative bacterial plant pathogens. (9-13).

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research, 32*(5), 1792-1797.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics, 26*(19), 2460-2461.

Finkmann, W., Altendorf, K., Stackebrandt, E., & Lipski, A. (2000). Characterization of N2O-producing Xanthomonas-like isolates from biofilters as Stenotrophomonas nitritireducens sp. nov., Luteimonas mephitis gen. nov., sp. nov. and Pseudoxanthomonas broegbernensis gen. nov., sp. nov. *International journal of systematic and evolutionary microbiology, 50*(1), 273-282.

Gautam, V., Patil, P. P., Bansal, K., Kumar, S., Kaur, A., Singh, A., . . . Patil, P. B. (2021). Description of Stenotrophomonas sepilia sp. nov., isolated from blood culture of a hospitalized patient as a new member of Stenotrophomonas maltophilia complex. *New Microbes New Infections*, 100920.

Gómez-Valero, L., Rocha, E. P., Latorre, A., & Silva, F. J. (2007). Reconstructing the ancestor of Mycobacterium leprae: the dynamics of gene loss and genome reduction. *Genome research, 17*(8), 1178-1185.

Hauben, L., Vauterin, L., Swings, J., & Moore, E. (1997). Comparison of 16S ribosomal DNA sequences of all Xanthomonas species. *International Journal of Systematic Evolutionary Microbiology, 47*(2), 328-335.

Hayashi Sant'Anna, F., Bach, E., Porto, R. Z., Guella, F., Hayashi Sant'Anna, E., & Passaglia, L. M. (2019). Genomic metrics made easy: what to do and where to go in the new era of bacterial taxonomy. *Critical reviews in microbiology*, 1-19.

He, Y.-W., Cao, X.-Q., & Poplawsky, A. R. (2020). Chemical Structure, Biological Roles, Biosynthesis and Regulation of the Yellow Xanthomonadin Pigments in the Phytopathogenic Genus Xanthomonas. *Molecular Plant-Microbe Interactions, 33*(5), 705-714.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., . . . Kuhn, M. (2015). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic acids research, 44*(D1), D286-D293.

Hugh, R., & Ryschenkow, E. (1961). Pseudomonas maltophilia, an alcaligenes-like species. *Microbiology, 26*(1), 123-132.

Indu, I., Ch, S., & Ch V, R. (2019). Taxogenomics resolves conflict in the genus Rhodobacter: a two and half decades pending thought to reclassify the genus Rhodobacter. *Frontiers in microbiology, 10*, 2480.

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature communications, 9*(1), 5114.

Katzen, F., Ferreiro, D. U., Oddo, C. G., Ielmini, M. V., Becker, A., Pühler, A., & Ielpi, L. (1998). Xanthomonas campestris pv. campestrisgum Mutants: Effects on Xanthan Biosynthesis and Plant Virulence. *Journal of bacteriology, 180*(7), 1607-1617.

Konstantinidis, K. T., & Tiedje, J. M. (2005). Towards a genome-based taxonomy for prokaryotes. *Journal of bacteriology, 187*(18), 6258-6264.

Kumar, S., Bansal, K., Patil, P. P., Kaur, A., Kaur, S., Jaswal, V., . . . Patil, P. B. (2020). Genomic insights into evolution of extensive drug resistance in Stenotrophomonas maltophilia complex. *Genomics, 112*(6), 4171-4178.

Kumar, S., Bansal, K., Patil, P. P., & Patil, P. B. (2019). Phylogenomics insights into order and families of Lysobacterales. *Access Microbiology, 1*(2).

Lee, D. S., Ryu, S. H., Hwang, H. W., Kim, Y.-J., Park, M., Lee, J. R., . . . Jeon, C. O. (2008). Pseudoxanthomonas sacheonensis sp. nov., isolated from BTEX-contaminated soil in Korea, transfer of Stenotrophomonas dokdonensis Yoon et al. 2006 to the genus Pseudoxanthomonas as Pseudoxanthomonas dokdonensis comb. nov. and emended description of the genus Pseudoxanthomonas. *International Journal of Systematic Evolutionary Microbiology, 58*(9), 2235-2240.

Liao, H., Lin, X., Li, Y., Qu, M., & Tian, Y. J. M. (2020). Reclassification of the taxonomic framework of orders cellvibrionales, oceanospirillales, pseudomonadales, and alteromonadales in class gammaproteobacteria through phylogenomic tree analysis. *Msystems, 5*(5), e00543-00520.

Lu, H., Patil, P., Van Sluys, M.-A., White, F. F., Ryan, R. P., Dow, J. M., . . . Sonti, R. (2008). Acquisition and evolution of plant pathogenesis–associated gene clusters and candidate determinants of tissue-specificity in Xanthomonas. *PLoS One, 3*(11), e3828.

Margos, G., Fingerle, V., Cutler, S., Gofton, A., Stevenson, B., & Estrada-Peña, A. (2019). Controversies in bacterial taxonomy: The example of the genus Borrelia. *Ticks and Tick-borne Diseases*, 101335.

Monteiro-Vitorello, C. B., De Oliveira, M. C., Zerillo, M. M., Varani, A. M., Civerolo, E., & Sluys, M.-A. V. (2005). Xylella and Xanthomonas mobil'omics. *Omics: a journal of integrative biology, 9*(2), 146-159.

Moreira, L. M., de Souza, R. F., Almeida Jr, N. F., Setubal, J. C., Oliveira, J. C. F., Furlan, L. R., . . . da Silva, A. C. (2004). Comparative genomics analyses of citrus-associated bacteria. *Annu. Rev. Phytopathol., 42*, 163-184.

Mostowy, R., Croucher, N. J., Andam, C. P., Corander, J., Hanage, W. P., & Marttinen, P. (2017). Efficient inference of recent and ancestral recombination within bacterial populations. *Molecular biology evolution, 34*(5), 1167-1182.

Nakamura, T., Yamada, K. D., Tomii, K., & Katoh, K. (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics, 34*(14), 2490-2492.

Naushad, H. S., & Gupta, R. S. (2013). Phylogenomics and molecular signatures for species from the plant pathogen-containing order xanthomonadales. *PLoS One, 8*(2), e55216.

Naushad, S., Adeolu, M., Wong, S., Sohail, M., Schellhorn, H. E., & Gupta, R. S. (2015). A phylogenomic and molecular marker based taxonomic framework for the order Xanthomonadales: proposal to transfer the families Algiphilaceae and Solimonadaceae to the order Nevskiales ord. nov. and to create a new family within the order Xanthomonadales, the family Rhodanobacteraceae fam. nov., containing the genus Rhodanobacter and its closest relatives. *Antonie van Leeuwenhoek, 107*(2), 467-485.

Orata, F. D., Meier-Kolthoff, J. P., Sauvageau, D., & Stein, L. Y. (2018). Phylogenomic analysis of the gammaproteobacterial methanotrophs (Order Methylococcales) calls for the reclassification of members at the genus and species levels. *Frontiers in microbiology, 9*, 3162.

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

Palleroni, N. J., & Bradbury, J. F. (1993). Stenotrophomonas, a New Bacterial Genus for Xanthomonas maltophilia (Hugh 1980) Swings et al. 1983. *International journal of systematic and evolutionary microbiology, 43*(3), 606-609. doi:doi:10.1099/00207713-43-3-606

Pammel, L. H. (2017). Bacteriosis of Rutabaga. *Bulletin, 3*(27), 7.

Pandey, A., & Sonti, R. V. (2010). Role of the FeoB protein and siderophore in promoting virulence of Xanthomonas oryzae pv. oryzae on rice. *Journal of bacteriology, 192*(12), 3187-3203.

Parkinson, N., Aritua, V., Heeney, J., Cowie, C., Bew, J., & Stead, D. (2007). Phylogenetic analysis of Xanthomonas species by comparison of partial gyrase B gene sequences. *International Journal of Systematic Evolutionary Microbiology, 57*(12), 2881-2887.

Parkinson, N., Cowie, C., Heeney, J., & Stead, D. (2009). Phylogenetic structure of Xanthomonas determined by comparison of gyrB sequences. *International Journal of Systematic Evolutionary Microbiology, 59*(2), 264-274.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research, 25*(7), 1043-1055.

Patil, P. P., Kumar, S., Kaur, A., Midha, S., Bansal, K., & Patil, P. B. (2021). Global transcriptome analysis of Stenotrophomonas maltophilia in response to growth at human body temperature. *Microbial genomics, 7*(7). doi:https://doi.org/10.1099/mgen.0.000600

Patil, P. P., Kumar, S., Midha, S., Gautam, V., & Patil, P. B. (2018). Taxonogenomics reveal multiple novel genomospecies associated with clinical isolates of Stenotrophomonas maltophilia. *Microbial genomics, 4*(8).

Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cociancich, S., . . . Jacques, M.-A. (2009). The complete genome sequence of Xanthomonas albilineans provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC genomics, 10*(1), 1-15.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One, 5*(3).

Qin, Q.-L., Xie, B.-B., Zhang, X.-Y., Chen, X.-L., Zhou, B.-C., Zhou, J., . . . Zhang, Y.-Z. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *Journal of bacteriology, 196*(12), 2210-2215.

Rajagopal, L., Sundari, C. S., Balasubramanian, D., & Sonti, R. V. (1997). The bacterial pigment xanthomonadin offers protection against photodamage. *FEBS letters, 415*(2), 125-128.

Ramírez-Gómez, X. S., Jiménez-García, S. N., Campos, V. B., & Campos, M. L. G. (2019). Plant Metabolites in Plant Defense Against Pathogens. In *Plant Pathology and Management of Plant Diseases*: IntechOpen.

Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences, 106*(45), 19126-19131.

Rodriguez-R, L. M., Grajales, A., Arrieta-Ortiz, M. L., Salazar, C., Restrepo, S., & Bernal, A. (2012). Genomes-based phylogeny of the genus Xanthomonas. *BMC microbiology, 12*(1), 1-14.

Ryan, R. P., Vorhölter, F.-J., Potnis, N., Jones, J. B., Van Sluys, M.-A., Bogdanove, A. J., & Dow, J. M. (2011). Pathogenomics of Xanthomonas: understanding bacterium–plant interactions. *Nature Reviews Microbiology, 9*(5), 344-355.

Saddler, G. S., & Bradbury, J. F. (2005). Xanthomonadalesord. nov. In *Bergey's manual® of systematic bacteriology* (pp. 63-122): Springer.

Salvetti, E., Harris, H. M., Felis, G. E., & O'Toole, P. W. (2018). Comparative genomics of the genus Lactobacillus reveals robust phylogroups that provide the basis for reclassification. *Appl. Environ. Microbiol., 84*(17), e00993-00918.

Sangal, V., Goodfellow, M., Jones, A. L., Schwalbe, E. C., Blom, J., Hoskisson, P. A., & Sutcliffe, I. C. (2016). Next-generation systematics: an innovative approach to resolve the structure of complex prokaryotic taxa. *Scientific reports, 6*(1), 1-12.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics, 30*(14), 2068-2069.

Studholme, D. J., Wasukira, A., Paszkiewicz, K., Aritua, V., Thwaites, R., Smith, J., & Grant, M. (2011). Draft genome sequences of Xanthomonas sacchari and two banana-associated xanthomonads reveal insights into the Xanthomonas group 1 clade. *Genes, 2*(4), 1050-1065.

Swings, J., De Vos, P., den MOOTER, M. V., & De Ley, J. (1983). Transfer of Pseudomonas maltophilia Hugh 1981 to the Genus Xanthomonas as Xanthomonas maltophilia (Hugh 1981) comb. nov. *International Journal of Systematic Evolutionary Microbiology, 33*(2), 409-413.

Van Sluys, M., Monteiro-Vitorello, C., Camargo, L., Menck, C., Da Silva, A., Ferro, J., . . . Simpson, A. (2002). Comparative genomic analysis of plant-associated bacteria. *Annual review of phytopathology, 40*(1), 169-189.

Wells, J. M., Raju, B. C., Hung, H.-Y., Weisburg, W. G., Mandelco-Paul, L., & Brenner, D. J. (1987). Xylella fastidiosa gen. nov., sp. nov: Gram-Negative, Xylem-Limited, Fastidious Plant Bacteria Related to Xanthomonas spp. *International journal of systematic and evolutionary microbiology, 37*(2), 136-143. doi:doi:10.1099/00207713-37-2-136

Yang, P., Vauterin, L., Vancanneyt, M., Swings, J., & Kersters, K. (1993). Application of fatty acid methyl esters for the taxonomic analysis of the genus Xanthomonas. *Systematic Applied Microbiology, 16*(1), 47-71.

Yarza, P., Ludwig, W., Euzéby, J., Amann, R., Schleifer, K.-H., Glöckner, F. O., & Rosselló-Móra, R. (2010). Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. *Systematic Applied Microbiology, 33*(6), 291-299.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., . . . Glöckner, F. O. (2014). The SILVA and "all-species living tree project (LTP)" taxonomic frameworks. *Nucleic acids research, 42*(D1), D643-D648.

Zheng, J., Wittouck, S., Salvetti, E., Franz, C. M., Harris, H. M., Mattarelli, P., . . . Walter, J. (2020). A taxonomic note on the genus Lactobacillus: Description of 23 novel genera, emended description of the genus Lactobacillus Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. *International Journal of Systematic Evolutionary Microbiology, 70*(4), 2782-2858.

Figures and Tables legends

Figure 1: Phylogenetic construction for XSXP phylogroup. a. Both the mega species groups are represented by blue boxes and four species complexes are represented by colored boxes. *Luteimonas mephitis* DSM12574^{T} was used as an outgroup and bootstrap values are mentioned on the nodes. **b.** Both the mega species groups are represented by blue boxes and four species complexes are represented by colored boxes. *Luteimonas mephitis* DSM12574^{T} was used as an outgroup and bootstrap vales are mentioned on the nodes.

Figure 2: Heatmap of genome similarity for XSXP genomes. a. Matrix showing average amino acid identity (AAI) values amongst XSXP phylogroup**.** All the species complexes are

represented in the colored boxes green: clade I; blue: clade II; yellow: clade III and red: clade IV. **b.** Heatmap showing percentage of conserved proteins (POCP) amongst XSXP group**.** All the species complexes are represented in the colored boxes green: clade I; blue: clade II; yellow: clade III and red: clade IV.

Figure 3: Distribution of COG-based functional categories of unique genes across several group. (A) clade I, (B) clade III, (C) clade II and (D) clade IV. *Stenotrophomonas maltophilia* complex represents the gene count across the several COG classes.

Figure 4: Number of recombination events detected in XSXP phylogroup.

Figure 5: A) Recombination network generated by HERO. Outer ring represents sequence clusters and length of each fragment is proportional to number of recombination events affecting the SC. Ribbons connecting clusters that share recombination events and its thickness is proportional to the number of shared events. Color of the ribbon matches the donor cluster. B) Recombination network highlighting highways of recombination and non-highways are colored grey.

Figure 6: Heatmap showing status of *gum* gene cluster (upper panel) and xanthomonadin pigment (lower panel) in XSXP phylogroup. Here absence of the gene is denoted by yellow box and presence by black or green box. All the species complexes are represented in the colored boxes green: Clade I; blue: clade II; yellow: clade III and red: clade IV.

Table 1: Metadata of the strains used in the present study. Following data was obtained from list of prokaryotic names with standing in nomenclature (http://www.bacterio.net/) and NCBI (https://www.ncbi.nlm.nih.gov/). Star marked columns are obtained in the present study. The data which is not available is marked with - across Host taxonomy, Lifestyle, Location and Isolation year columns.

Supplementary figure 1: Core genome phylogeny of the XSXP phylogroup. Here, core genome alignment is obtained by using PIRATE.

Supplementary figure 2: OrthoANI value calculated for the XSXP phylogroup.

Supplementary table 1: Average nucleotide identity (ANI) and alignment fraction (AF) amongst the XSXP phylogroup.

Supplementary table 2: Genes unique to clade I, its COG classification and its protein sequence.

Supplementary table 3: Genes unique to clade III, its COG classification and its protein sequence.

Supplementary table 4: Genes unique to clade II, its COG classification and its protein sequence.

Supplementary table 5: Genes unique to clade IV, its COG classification and its protein sequence.

Supplementary table 6: Recombination events detected within XSXP phylogroup.

Supplementary table 7: Sequence clusters among the XSXP phylogroup.

Supplementary table 8: Genes undergoing recombination amongst the XSXP pangenome.

Table 1: Metadata of the strains used in the present study. Following data was obtained from list of prokaryotic names with standing in nomenclature (http://www.bacterio.net/) and NCBI (https://www.ncbi.nlm.nih.gov/). Star marked columns are obtained in the present study. The data which is not available is marked with - across Host taxonomy, Lifestyle, Location and Isolation year columns.

Strain	Genome size	CDS	Genome status	Source/Host	Host taxonomy	Lifestyle	Location	Isolation year	GC%	# tRNA	Completeness*/ Contaminations*
Xanthomonas vasicola NCPPB 2417^T	4.9	3960	Draft	*Sorghum vulgare*	Tracheophytes, Angiosperms, Monocots, Commelinids, Poales, Poaceae	Pathogenic	New Zealand	1969	63.3	51	100/0.07
Xanthomonas oryzae ATCC 35933^T	4.2	3385	Draft	*Oryza sativa* (leaf blight and wilt)	Tracheophytes, Angiosperms, Monocots, Commelinids, Poales, Poaceae	Pathogenic	-	-	63.9	49	100/0
Xanthomonas euvesicatoria 85-10	5.1	4416	Complete	Infected pepper plant	Tracheophytes, Angiosperms, Eudicots, Asterids, Solanales, Solanaceae	Pathogenic	-	-	64.7	54	99.64/0
Xanthomonas alfalfae LMG 495^T	5.0	3937	Draft	*Medicago sativa*	Tracheophytes, Angiosperms, Eudicots, Rosids, Fabales, Fabaceae	Pathogenic	Belgium: Merelbeke	2014	64.8	44	98.79/2.99
Xanthomonas perforans CFBP 7293^T	5.0	4206	Draft	Leaves of *Solanum lycopersicum*	Tracheophytes, Angiosperms, Eudicots, Asterids, Solanales, Solanaceae	Pathogenic	USA: Florida	-	64.9	54	99.64/0.2
Xanthomonas phaseoli CFBP 412	5.0	4175	Complete	*Phaseolus vulgaris*	Tracheophytes, Angiosperms, Eudicots, Rosids, Fabales, Fabaceae	Pathogenic	USA	-	65	54	99.64/0.24
Xanthomonas citri LMG 9322^T	5.1	4355	Draft	*Citrus aurantifolia*	Tracheophytes, Angiosperms, Eudicots, Rosids, Sapindales, Rutaceae	Pathogenic	USA: Florida	1915	64.6	50	98.19/1.09
Xanthomonas axonopodis DSM 3585^T	4.4	3263	Draft	*Axonopus scoparius*	Tracheophytes, Angiosperms, Monocots, Commelinids, Poales, Poaceae	Pathogenic	Colombia	1949	64.5	50	99.52/0.12
Xanthomonas prunicola CFBP 8353^T	5.3	4208	Draft	*Prunus persica* var. nectarina	Tracheophytes, Angiosperms, Eudicots, Rosids, Rosales, Rosaceae	Pathogenic	Spain: Murcia, Abaran	2015	64	53	100/0
Xanthomonas bromi CFBP 1976^T	4.9	3845	Draft	*Bromus carinatus*	Tracheophytes, Angiosperms, Monocots, Commelinids, Poales, Poaceae	Pathogenic	France	1980	64	53	100/0
Xanthomonas nasturtii WHRI 8853^T	4.8	3981	Draft	Watercress	Tracheophytes, Angiosperms, Eudicots, Rosids, Brassicales, Brassicaceae	Pathogenic	USA: Florida	2014	64.5	56	100/0.36
Xanthomonas cynarae CFBP 4188^T	5.0	4102	Draft	*Cynara scolymus*	Tracheophytes, Angiosperms, Eudicots, Asterids, Asterales, Asteraceae	Pathogenic	France: Bretagne	1996	63.7	51	100/0.41
Xanthomonas gardneri ICMP 7383	5.3	4523	Complete	Tomato field	Tracheophytes, Angiosperms, Eudicots, Asterids, Solanales, Solanaceae	Pathogenic	New Zealand	1980	63.5	53	99.64/0.68

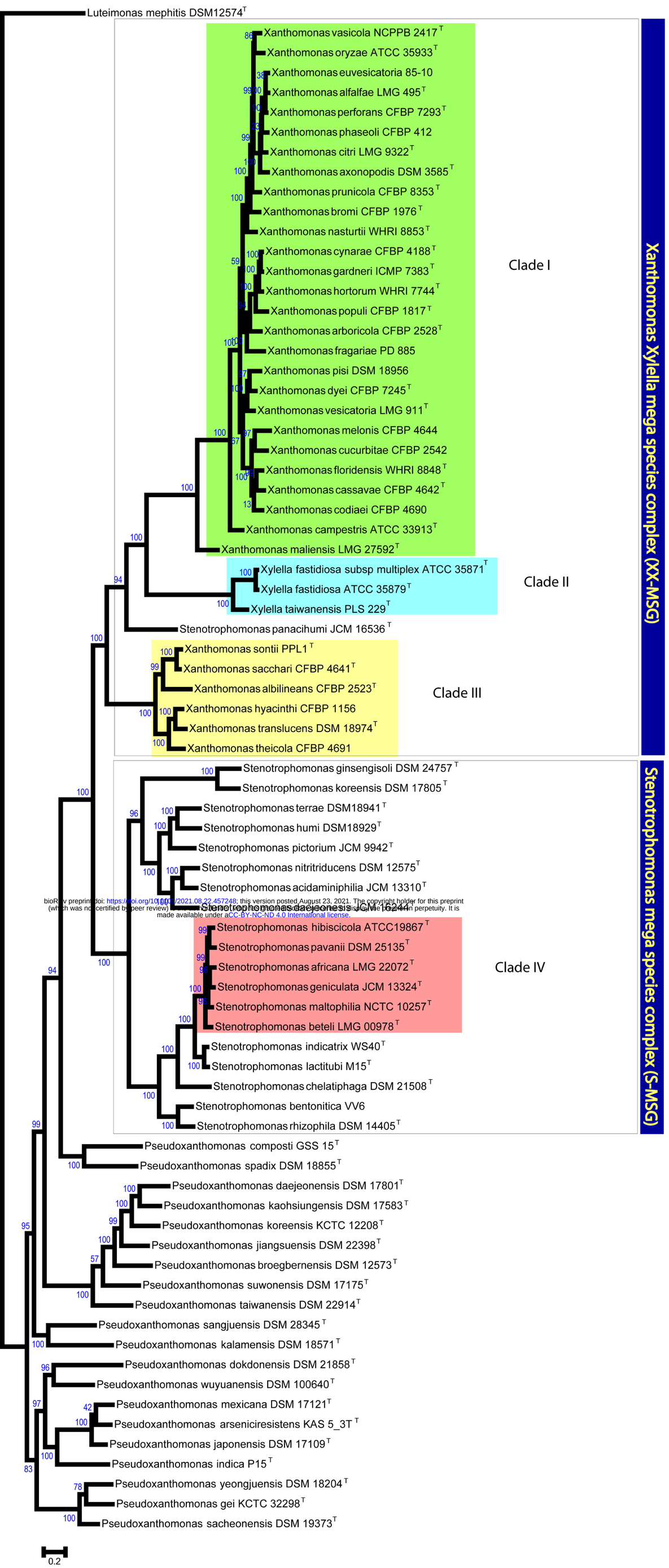
Xanthomonas hortorum WHRI 7744^{T}	5.5	4427	Draft	*Hedera helix*	Tracheophytes, Angiosperms, Eudicots, Asterids, Apiales, Araliaceae	Pathogenic	USA	1961	63.6	54	100/1.03
Xanthomonas populi CFBP 1817^{T}	4.3	3222	Draft	*Populus x canadensis* cv. regenerata)	Tracheophytes, Angiosperms, Eudicots, Rosids, Malpighiales, Salicaceae	Pathogenic	Oise, Noyon	1957	63.3	48	99.64/0.14
Xanthomonas arboricola CFBP 2528^{T}	5.0	4124	Draft	Twig from a walnut tree	Tracheophytes, Angiosperms, Eudicots, Asterids, Fagales, Juglandaceae	Pathogenic	New Zealand	1956	65.5	51	100/0.41
Xanthomonas fragariae PD 885	4.2	3263	Complete	Diseased strawberry	Tracheophytes, Angiosperms, Eudicots, Rosids, Rosales, Rosaceae	Pathogenic	-	-	62.3	52	99.64/0.36
Xanthomonas pisi DSM 18956	4.2	3385	Draft	*Pisum sativum*; Leaf spot	Tracheophytes, Angiosperms, Eudicots, Rosids, Fabales, Fabaceae	Pathogenic	Japan	1997	64.7	43	80.23/0
Xanthomonas dyei CFBP 7245^{T}	5.3	4278	Draft	*Metrosideros excelsa*; Angular leaf lesion	Tracheophytes, Angiosperms, Eudicots, Rosids, Myrtales, Myrtaceae	Pathogenic	Bay of Plenty Omahanui	1993	64.3	51	100/0.36
Xanthomonas vesicatoria LMG 911^{T}	5.1	4343	Complete	*Lycopersicon esculentum*	Tracheophytes, Angiosperms, Eudicots, Asterids, Solanales, Solanaceae	Pathogenic	New Zealand	1955	64.3	55	100/0.36
Xanthomonas melonis CFBP 4644	4.6	3872	Draft	*Cucumis melo*	Tracheophytes, Angiosperms, Eudicots, Rosids, Cucurbitales, Cucurbitaceae	Pathogenic	Brazil	1974	66.1	52	99.64/0.18
Xanthomonas cucurbitae CFBP 2542^{T}	4.4	3553	Draft	*Cucurbita maxima*	Tracheophytes, Angiosperms, Eudicots, Rosids, Cucurbitales, Cucurbitaceae	Pathogenic	New Zealand	1968	65.4	51	100/0.36
Xanthomonas floridensis WHRI 8848^{T}	5.2	4286	Draft	Watercress	Tracheophytes, Angiosperms, Eudicots, Rosids, Brassicales, Brassicaceae	Pathogenic	USA: Florida	2014	65.4	52	100/0.36
Xanthomonas cassavae CFBP 4642^{T}	5.2	4218	Draft	*Manihot esculenta*	Tracheophytes, Angiosperms, Eudicots, Rosids, Malpighiales, Euphorbiaceae	Pathogenic	Malawi	1951	65.2	51	99.64/0.93
Xanthomonas codiaei CFBP 4690	5.0	4125	Draft	*Codiaeum variegatum*	Tracheophytes, Angiosperms, Eudicots, Rosids, Malpighiales, Euphorbiaceae	Pathogenic	USA: Florida	1987	66	53	99.28/0
Xanthomonas campestris pv. *campestris* ATCC 33913^{T}	5.0	4179	Complete	Plant (*Brassica oleracea* var. gemmifera)	Tracheophytes, Angiosperms,Eudicots, Rosids, Brassicales, Brassicaceae	Pathogenic	-	-	65.1	53	99.64/0
Xanthomonas maliensis LMG 27592^{T}	5.2	4197	Draft	*Oryza sativa*	Tracheophytes, Angiosperms, Monocots, Commelinids, Poales, Poaceae	Non-pathogenic	Mali: Africa	2009	66.2	53	96.74/0.3
Xylella taiwanensis PLS229^{T}	2.7	2047	Draft	Pear	Tracheophytes, Angiosperms, Eudicots, Rosids, Rosales, Rosaceae	Pathogenic	Taiwan	-	53	47	89.88/0.12
Xylella fastidiosa ATCC 35879^{T}	2.5	2105	Draft	Grapevine	Tracheophytes, Angiosperms, Eudicots, Rosids, Vitales, Vitaceae	Pathogenic	USA:Florida	1987	51.8	49	99.23/0
Stenotrophomonas panacihumi JCM 16536^{T}	3.9	3497	Draft	Soil of a ginseng field	-	Environmental	South korea	2010	68.8	51	97.59/0.34
Xanthomonas sontii PPL1^{T}	4.8	3937	Draft	*Oryza sativa*	Tracheophytes, Angiosperms, Monocots, Commelinids, Poales, Poaceae	Non-pathogenic	India	2012	69	51	96.31/0.22
Xanthomonas sacchari CFBP 4641^{T}	4.9	4098	Draft	*Saccharum officinarum*	Tracheophytes, Angiosperms, Monocots, Commelinids, Poales, Poaceae	Pathogenic	Guadeloupe	1980	69.1	51	100/0.41
Xanthomonas albilineans CFBP 2523^{T}	3.6	2967	Draft	*Saccharum officinarum*	Tracheophytes, Angiosperms, Monocots, Commelinids, Poales, Poaceae	Pathogenic	Fiji	1961	63.1	49	99.96/0.34

Xanthomonas hyacinthi CFBP 1156	4.9	4003	Draft	*Hyacinthus orientalis*	Tracheophytes, Angiosperms, Monocots, Asparagales, Asparagaceae	Pathogenic	Netherlands	1958	68.1	51	100/0.03
Xanthomonas translucens DSM 18974^T	4.4	3788	Draft	*Hordeum vulgare*	Tracheophytes, Angiosperms, Monocots, Commelinids, Poales, Poaceae	Pathogenic	-	-	67.7	53	98.5/0.12
Xanthomonas theicola CFBP 4691	4.5	3630	Draft	*Camellia sinensis*	Tracheophytes, Angiosperms, Eudicots, Asterids, Ericales, Theaceae	Pathogenic	Japan	1974	68.4	49	99.95/0
Stenotrophomonas ginsengisoli DSM 24757^T	3.3	3001	Draft	Soil from ginseng field	-	Environmental	South Korea	2010	65.9	76	98.72/0.76
Stenotrophomonas koreensis DSM 17805^T	3.0	2690	Draft	Compost	-	--	South Korea	2003	66.1	75	98.16/0.79
Stenotrophomonas terrae DSM 18941^T	4.4	3730	Draft	Soil	-	Environmental	Belgium	2007	63.9	60	99.19/1.38
Stenotrophomonas humi DSM 18929^T	4.1	3591	Draft	Soil	-	Environmental	Belgium	2007	64	62	99.6/1.03
Stenotrophomonas pictorum JCM 9942^T	3.3	3150	Draft	Soil	-	Environmental	-	1928	66	61	99.84/0.32
Stenotrophomonas nitritireducens DSM 12575^T	3.9	3476	Draft	Laboratory scale biofilters supplied with ammonia or dimethyl disulfide and ammonia	-	-	Germany	1997	68.3	54	95.5/0.78
Stenotrophomonas acidaminiphila JCM 13310^T	3.9	3511	Draft	Sludge from a lab-scale anaerobic chemical waste water reactor	-	Environmental	Mexico	1999	68.8	56	96.69/1.51
Stenotrophomonas daejeonensis JCM 16244^T	3.2	2888	Draft	Sewage water	-	Environmental	South Korea	2010	68.6	56	99.59/0.86
stenotrophomonas hibiscicola ATCC 19867^T	4.4	3928	Draft	Plant	-	-	-	-	66.4	64	100/0
Stenotrophomonas pavanii DSM 25135^T	4.3	3783	Draft	Stems of sugar cane	Tracheophytes, Angiosperms, Monocots, Commelinids, Poales, Poaceae	Symbiotic	Brazil	2011	66.2	67	100/0
Stenotrophomonas africana LMG 22072^T	4.5	3991	Draft	Cerebrospinal fluid	-	Opportunistic pathogen	Democratic Republic of the Congo	1994	66.3	65	97.05/0.23

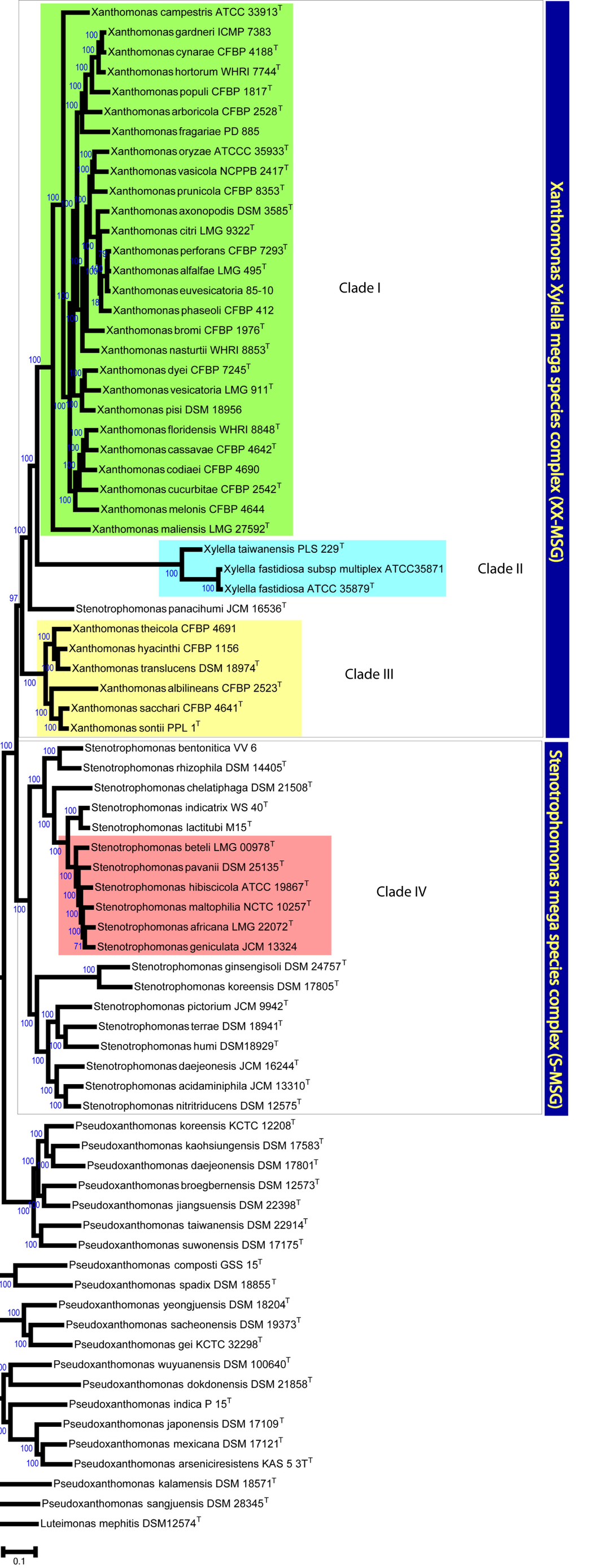
Pseudomonas geniculate JCM 13324^{T}	4.8	4339	Draft	Tap water	-	Environmental	USA	1895	66.2	66	99.76/0.41
Stenotrophomonas maltophilia NCTC 10257^{T}	5.0	4491	Complete	Mouth	-	Opportunistic pathogen	Germany	1900/1961	66.1	73	100/0.34
Stenotrophomonas beteli LMG 00978^{T}	4.4	3907	Draft	*Piper betle*	Tracheophytes, Angiosperms, Magnoliids, Piperales, Piperaceae	-	Sri Lanka	1928	66.8	67	99.08/0
Stenotrophomonas indicatrix WS40^{T}	4.5	4067	Draft	Dirty dishes	-	Environmental	Germany: Bonn	2013	66.4	65	99.91/0.57
Stenotrophomonas lactitubi M15^{T}	4.8	4378	Draft	Milking machine biofilm	-	-	Germany: Koenigswinter-Vinxel	2014	65.9	66	100/0.41
Stenotrophomonas chelatiphaga DSM 21508^{T}	3.9	3420	Draft	Municipal sewage sludge	-	Environmental	Russia	2009	66.9	62	98.75/0.04
Stenotrophomonas bentonitica VV6	4.3	3805	Draft	Arctic sea water near B02 station	-	Environmental	-	-	66.5	60	99.91/1.93
Stenotrophomonas rhizophila DSM 14405^{T}	4.6	3971	Draft	Root	-	Symbiotic	Germany: Rostock	-	67.3	67	100/0.21
Pseudoxanthomonas composti GSS 15^{T}	4.3	3664	Draft	Compost	-	Environmental	China	-	68.2	52	99.31/0.8
Pseudoxanthomonas spadix DSM 18855^{T}	3.3	3042	Draft	Oil-contaminated soil	-	Environmental	Taiwan	Before 2006	67.8	48	98.97/0.59
Pseudoxanthomonas daejeonensis DSM 17801^{T}	3.5	3143	Draft	Soil from a ginseng field	-	Environmental	-	-	68.8	58	99.66/0.11
Pseudoxanthomonas kaohsiungensis DSM 17583^{T}	3.7	3420	Draft	Oil-polluted site	-	Environmental	Taiwan	2003	69.7	52	99.66/1.42
Pseudoxanthomonas koreensis KCTC 12208^{T}	3.0	2681	Draft	Soil from a ginseng field	-	Environmental	South Korea	-	70.1	52	99.48/0.04
Pseudoxanthomonas jiangsuensis DSM 22398^{T}	3.7	3389	Draft	DDT-contaminated soil	-	Environmental	China	2008	70.3	50	99.31/1.03
Pseudoxanthomonas broegbernensis DSM 12573^{T}	3.5	3024	Draft	Ammonia-supplied biofilters	-	Environmental	Germany	Before 1998	70.6	54	99.66/1.74
Pseudoxanthomonas suwonensis DSM 17175^{T}	3.4	3071	Draft	Cotton waste composts	-	Environmental	South Korea	Before 2005	70.3	55	99.66/0.34

Pseudoxanthomonas taiwanensis DSM 22914^T	3.0	2729	Draft	Hot spring	-	Environmental	Taiwan	Before 2009	72.08	56	99.16/0.34
Pseudoxanthomonas sangjuensis DSM 28345^T	3.2	2908	Draft	Greenhouse soil	-	Environmental	Republic of Korea	Before 2014	68.6	51	99.95/0.46
Pseudoxanthomonas kalamensis DSM 18571^T	3.0	2697	Draft	Soil contaminated with polycyclic aromatic hydrocarbons & polychlorinated biphenyls	-	Environmental	United States of America	Before 2006	65.9	48	99.84/1.5
Pseudoxanthomonas dokdonensis DSM 21858^T	3.5	3153	Draft	Soil	-	Environmental	South Korea	Before 2008	64.4	50	99.59/0.41
Pseudoxanthomonas wuyuanensis DSM 100640^T	4.6	4003	Draft	Saline-alkaline soil	-	Environmental	China	Before 2015	65.7	50	100/1.08
Pseudoxanthomonas mexicana DSM 17121^T	3.9	3660	Draft	Anaerobic digester	-	Environmental	Mexico	Before 2005	67.4	54	99.49/1.71
Pseudoxanthomonas arseniciresistens KAS 5_3^T	3.9	4049	Draft	Arsenic contaminated ground water	-	Environmental	India	-	66.5	52	99.95/1.19
Pseudoxanthomonas japonensis DSM 17109^T	4.0	3673	draft	Urban riverside soil		Environmental	Japan	Before 2005	67.3	50	99.95/0.76
Pseudoxanthomonas indica P15^T	3.9	3593	Draft	Open hexachlorocyclohexane dumpsite soil	-	Environmental	India	-	65.4	49	99.89/0.47
Pseudoxanthomonas yeongjuensis DSM 18204^T	3.9	3389	Draft	Soil of a ginseng field	-	Environmental	Republic of Korea	2005	65.1	53	99.95/0.76
Pseudoxanthomonas gei KCTC 32298^T	3.4	3122	Draft	Stem of *Geum aleppicum* Jacq	Tracheophytes, Angiosperms, Eudicots, Rosids, Rosales, Rosaceae	Endophyte	Taibai Mountain in Shaanxi Province, north-west China	-	65.4	45	99.95/0.41

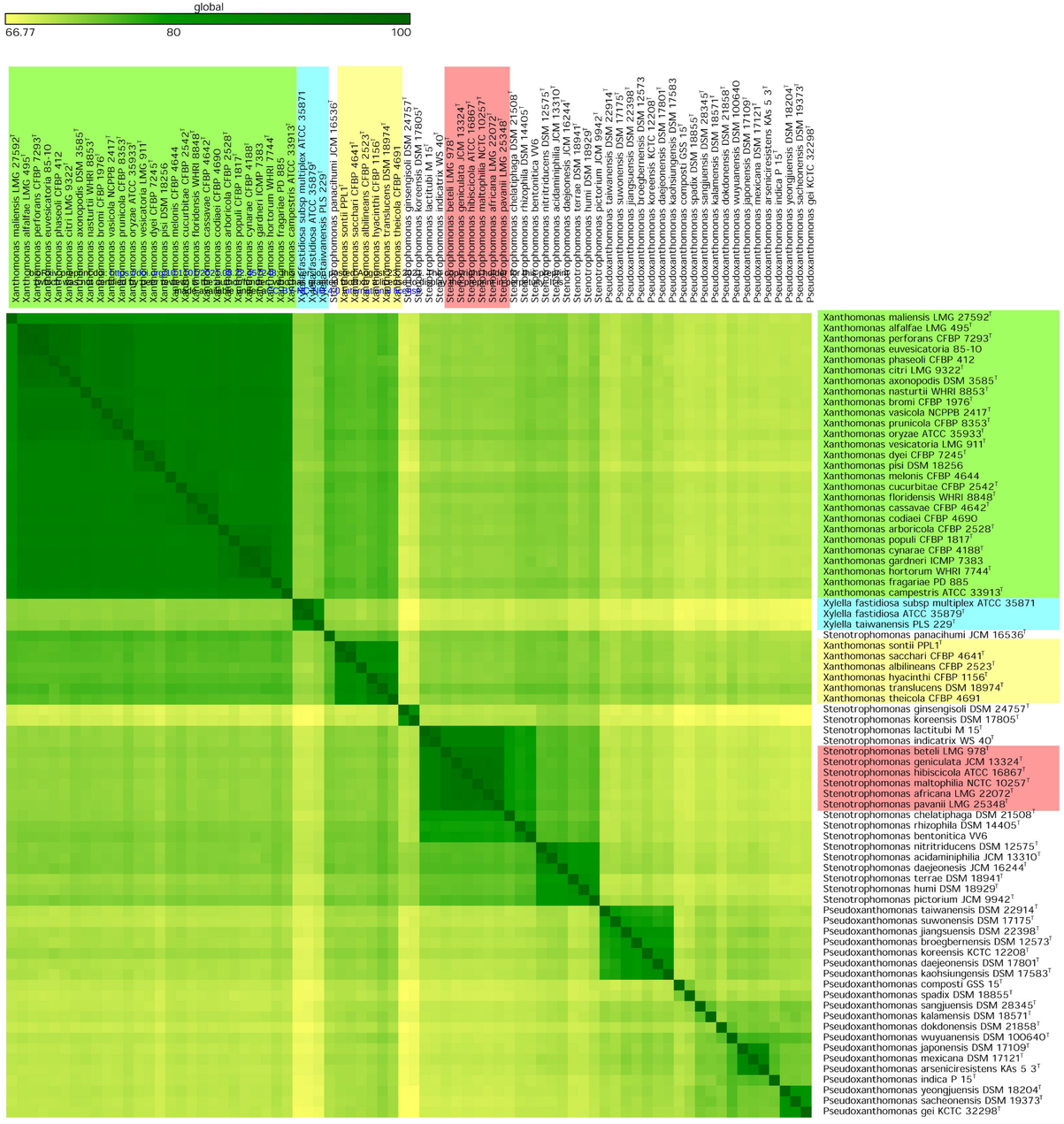
Pseudoxanthomonas sacheonensis DSM 19373^{T}	4.0	3582	Draft	BTEX-contaminated soil	-	Environmental	Republic of Korea	Before 2007	64.3	50	100/0.91



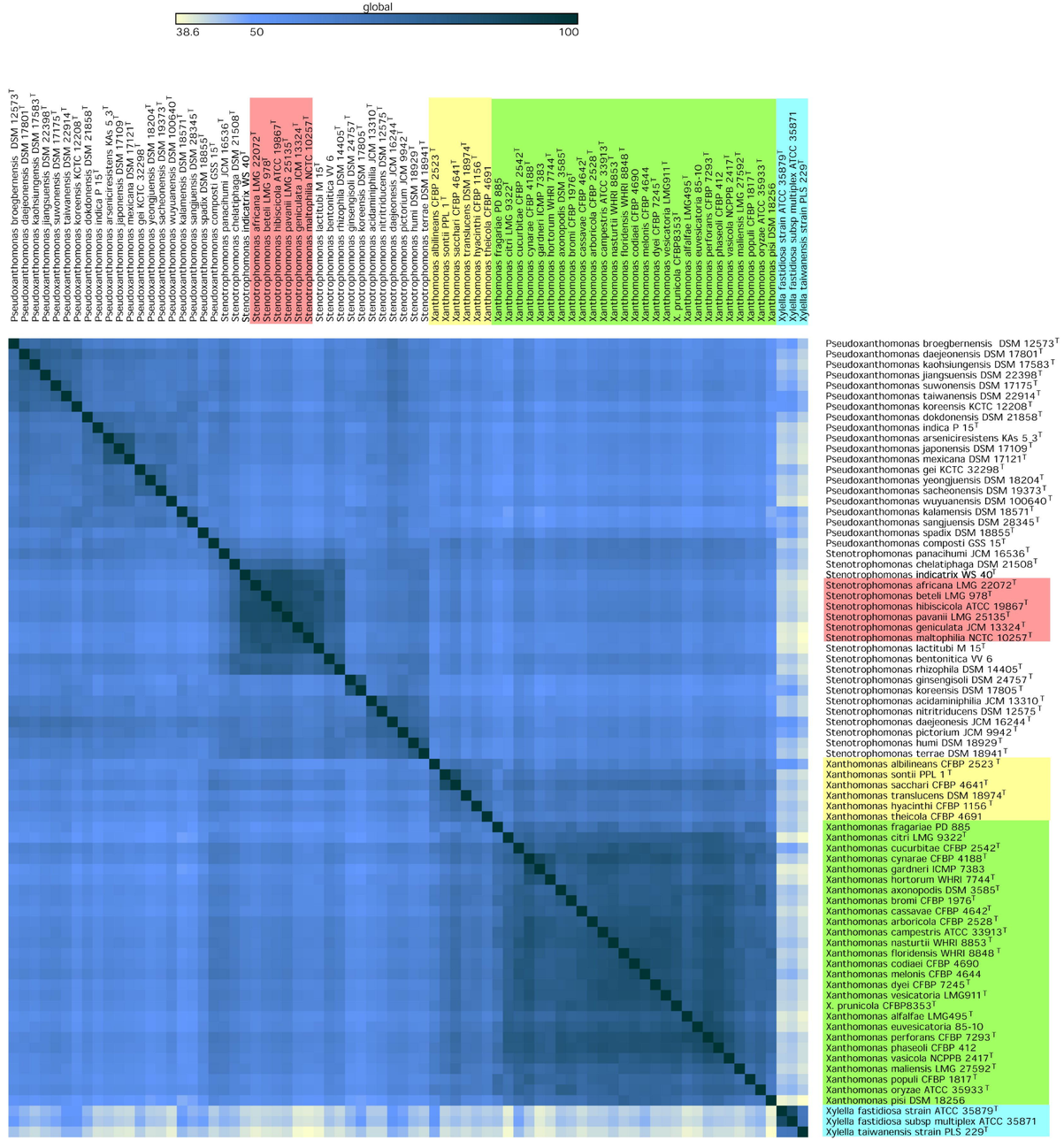
(A)



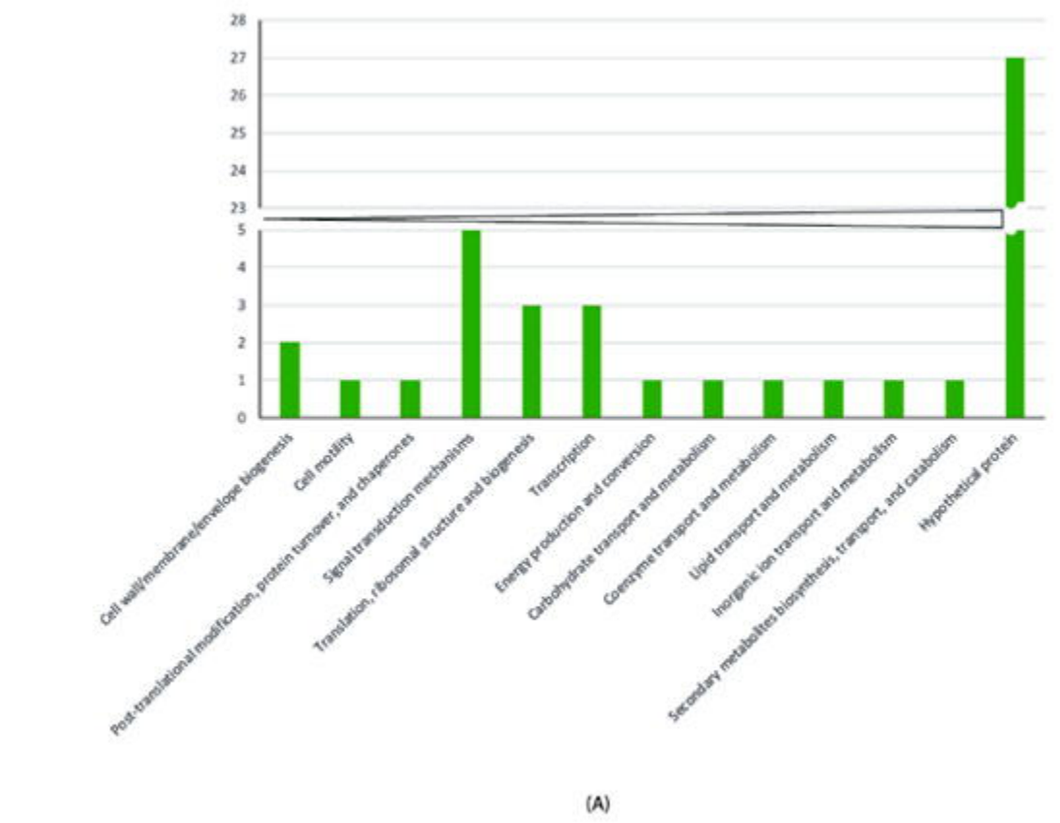
(B)


global
66.77
80
100
Xanthomonas maliensis LMG 27592T
Xanthomonas alfalfae LMG 495T
Xanthomonas perforans CFBP 7293T
Xanthomonas euvesicatoria 85-10
Xanthomonas phaseoli CFBP 412
Xanthomonas citri LMG 9322T
Xanthomonas axonopodis DSM 3585T
Xanthomonas nasturtii WHRI 8853T
Xanthomonas bromi CFBP 1976T
Xanthomonas vasicola NCPPB 2417T
Xanthomonas prunicola CFBP 8353T
Xanthomonas oryzae ATCC 35933T
Xanthomonas vesicatoria LMG 911T
Xanthomonas dyei CFBP 7245T
Xanthomonas pisi DSM 18256
Xanthomonas melonis CFBP 4644
Xanthomonas cucurbitae CFBP 2542T
Xanthomonas floridensis WHRI 8848T
Xanthomonas cassavae CFBP 4642T
Xanthomonas codiaei CFBP 4690
Xanthomonas arboricola CFBP 2528T
Xanthomonas populi CFBP 1817T
Xanthomonas cynarae CFBP 4188T
Xanthomonas gardneri ICMP 7383
Xanthomonas hortorum WHRI 7744T
Xanthomonas fragariae PD 885
Xanthomonas campestris ATCC 33913T
Xylella fastidiosa subsp multiplex ATCC 35871
Xylella fastidiosa ATCC 35879T
Xylella taiwanensis PLS 229T
Stenotrophomonas panachiumi JCM 16536T
Xanthomonas sontii PPL1T
Xanthomonas sacchari CFBP 4641T
Xanthomonas albilineans CFBP 2523T
Xanthomonas hyacinthi CFBP 1156T
Xanthomonas translucens DSM 18974T
Xanthomonas theicola CFBP 4691
Stenotrophomonas ginsengisoli DSM 24757T
Stenotrophomonas koreensis DSM 17805T
Stenotrophomonas lactitubi M 15T
Stenotrophomonas indicatrix WS 40T
Stenotrophomonas beteli LMG 978T
Stenotrophomonas geniculata JCM 13324T
Stenotrophomonas hibiscicola ATCC 16867T
Stenotrophomonas maltophilia NCTC 10257T
Stenotrophomonas pavanii LMG 25348T
Stenotrophomonas chelatiphaga DSM 21508T
Stenotrophomonas rhizophila DSM 14405T
Stenotrophomonas bentonitica VV6
Stenotrophomonas nitritriducens DSM 12575T
Stenotrophomonas acidaminiphilia JCM 13310T
Stenotrophomonas daejeonesis JCM 16244T
Stenotrophomonas terrae DSM 18941T
Stenotrophomonas humi DSM 18929T
Stenotrophomonas pictorium JCM 9942T
Pseudoxanthomonas taiwanensis DSM 22914T
Pseudoxanthomonas suwonensis DSM 17175T
Pseudoxanthomonas jiangsuensis DSM 22398T
Pseudoxanthomonas broegbernensis DSM 12573T
Pseudoxanthomonas koreensis KCTC 12208T
Pseudoxanthomonas daejeonensis DSM 17801T
Pseudoxanthomonas kaohsiungensis DSM 17583T
Pseudoxanthomonas composti GSS 15T
Pseudoxanthomonas spadix DSM 18855T
Pseudoxanthomonas sangjuensis DSM 28345T
Pseudoxanthomonas kalamensis DSM 18571T
Pseudoxanthomonas dokdonensis DSM 21858T
Pseudoxanthomonas wuyuanensis DSM 100640T
Pseudoxanthomonas japonensis DSM 17109T
Pseudoxanthomonas mexicana DSM 17121T
Pseudoxanthomonas arseniciresistens KAs 5_3T
Pseudoxanthomonas indica P 15T
Pseudoxanthomonas yeongjuensis DSM 18204T
Pseudoxanthomonas sacheonensis DSM 19373T
Pseudoxanthomonas gei KCTC 32298T

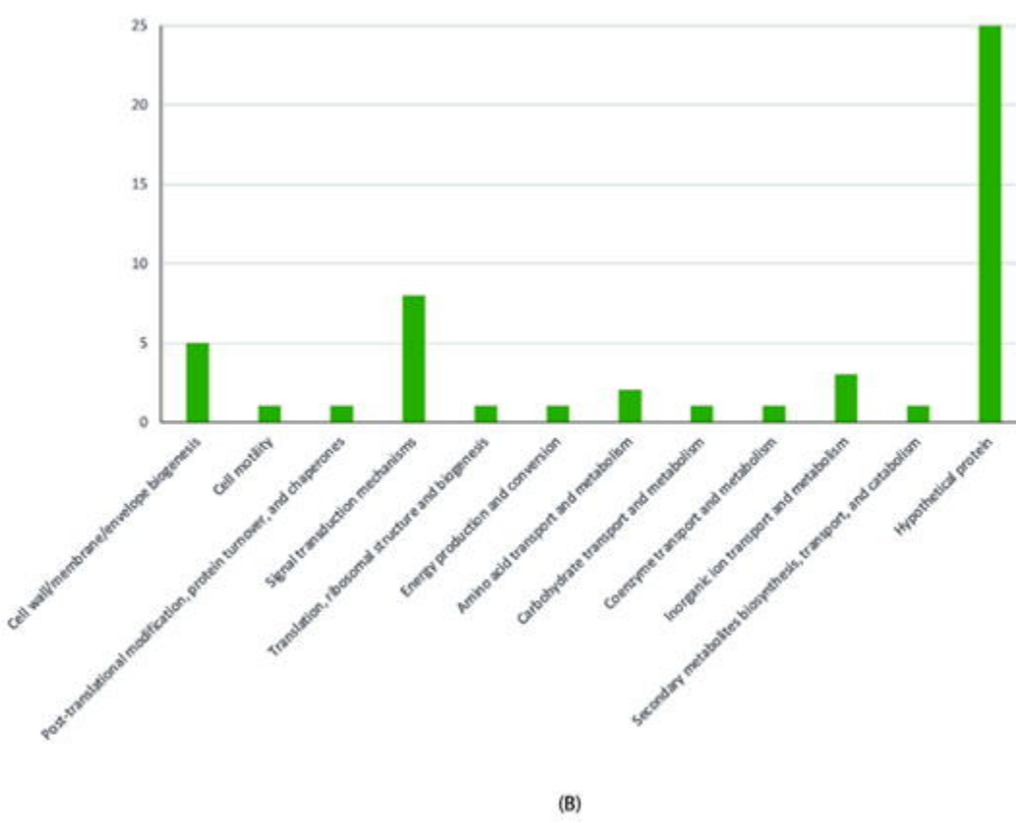
(A)


global
38.6
50
100
Pseudoxanthomonas broegbernensis DSM 12573T
Pseudoxanthomonas daejeonensis DSM 17801T
Pseudoxanthomonas kaohsiungensis DSM 17583T
Pseudoxanthomonas jiangsuensis DSM 22398T
Pseudoxanthomonas suwonensis DSM 17175T
Pseudoxanthomonas taiwanensis DSM 22914T
Pseudoxanthomonas koreensis KCTC 12208T
Pseudoxanthomonas dokdonensis DSM 21858T
Pseudoxanthomonas indica P 15T
Pseudoxanthomonas arseniciresistens KAs 5_3T
Pseudoxanthomonas japonensis DSM 17109T
Pseudoxanthomonas mexicana DSM 17121T
Pseudoxanthomonas gei KCTC 32298T
Pseudoxanthomonas yeongjuensis DSM 18204T
Pseudoxanthomonas sacheonensis DSM 19373T
Pseudoxanthomonas wuyuanensis DSM 100640T
Pseudoxanthomonas kalamensis DSM 18571T
Pseudoxanthomonas sangjuensis DSM 28345T
Pseudoxanthomonas spadix DSM 18855T
Pseudoxanthomonas composti GSS 15T
Stenotrophomonas panachiumi JCM 16536T
Stenotrophomonas chelatiphaga DSM 21508T
Stenotrophomonas indicatrix WS 40T
Stenotrophomonas africana LMG 22072T
Stenotrophomonas beteli LMG 978T
Stenotrophomonas hibiscicola ATCC 19867T
Stenotrophomonas pavanii LMG 25135T
Stenotrophomonas geniculata JCM 13324T
Stenotrophomonas maltophilia NCTC 10257T
Stenotrophomonas lactitubi M 15T
Stenotrophomonas bentonitica VV 6
Stenotrophomonas rhizophila DSM 14405T
Stenotrophomonas ginsengisoli DSM 24757T
Stenotrophomonas koreensis DSM 17805T
Stenotrophomonas acidaminiphilia JCM 13310T
Stenotrophomonas nitritriducens DSM 12575T
Stenotrophomonas daejeonesis JCM 16244T
Stenotrophomonas pictorium JCM 9942T
Stenotrophomonas humi DSM 18929T
Stenotrophomonas terrae DSM 18941T
Xanthomonas albilineans CFBP 2523T
Xanthomonas sontii PPL 1T
Xanthomonas sacchari CFBP 4641T
Xanthomonas translucens DSM 18974T
Xanthomonas hyacinthi CFBP 1156T
Xanthomonas theicola CFBP 4691
Xanthomonas fragariae PD 885
Xanthomonas citri LMG 9322T
Xanthomonas cucurbitae CFBP 2542T
Xanthomonas cynarae CFBP 4188T
Xanthomonas gardneri ICMP 7383
Xanthomonas hortorum WHRI 7744T
Xanthomonas axonopodis DSM 3585T
Xanthomonas bromi CFBP 1976T
Xanthomonas cassavae CFBP 4642T
Xanthomonas arboricola CFBP 2528T
Xanthomonas campestris ATCC 33913T
Xanthomonas nasturtii WHRI 8853T
Xanthomonas floridensis WHRI 8848T
Xanthomonas codiaei CFBP 4690
Xanthomonas melonis CFBP 4644
Xanthomonas dyei CFBP 7245T
Xanthomonas vesicatoria LMG911T
X. prunicola CFBP8353T
Xanthomonas alfalfae LMG495T
Xanthomonas euvesicatoria 85-10
Xanthomonas perforans CFBP 7293T
Xanthomonas phaseoli CFBP 412
Xanthomonas vasicola NCPPB 2417T
Xanthomonas maliensis LMG 27592T
Xanthomonas populi CFBP 1817T
Xanthomonas oryzae ATCC 35933T
Xanthomonas pisi DSM 18256
Xylella fastidiosa strain ATCC 35879T
Xylella fastidiosa subsp multiplex ATCC 35871
Xylella taiwanensis strain PLS 229T

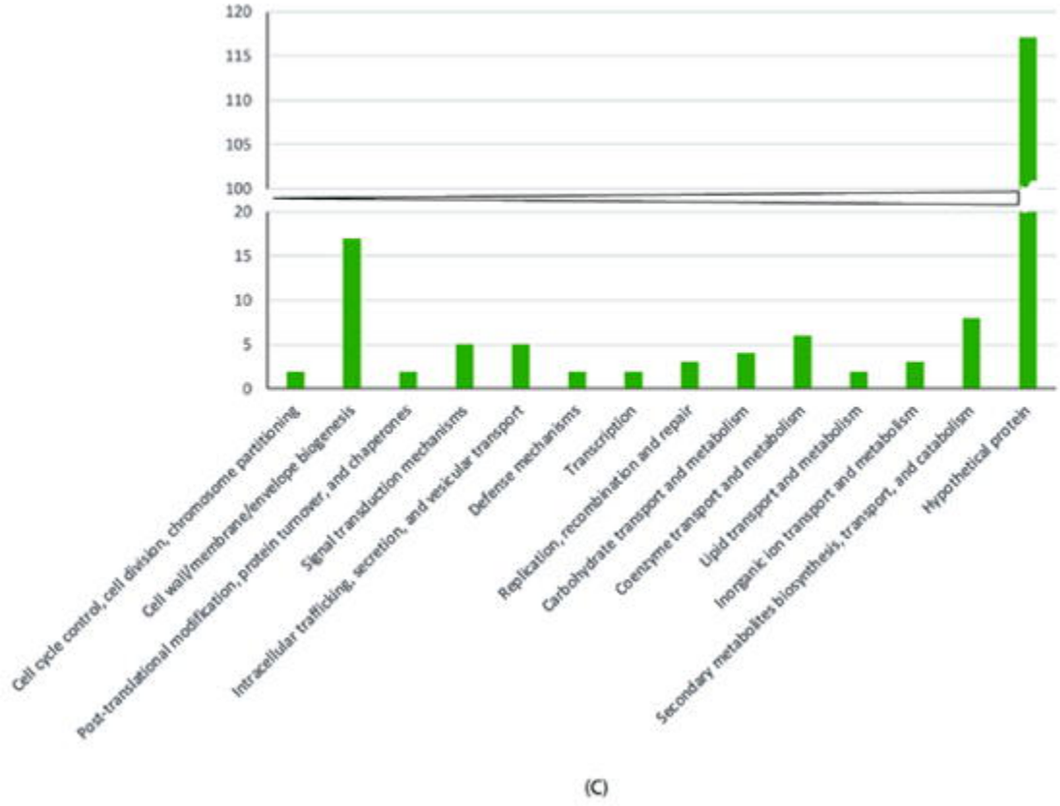
(B)


28
27
26
25
24
23
5
4
3
2
1
0
Cell wall/membrane/envelope biogenesis
Cell motility
Post-translational modification, protein turnover , and chaperones
Signal transduction mechanisms
Translation, ribosomal structure and biogenesis
Transcription
Energy production and conversion
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis, transport, and catabolism
Hypothetical protein

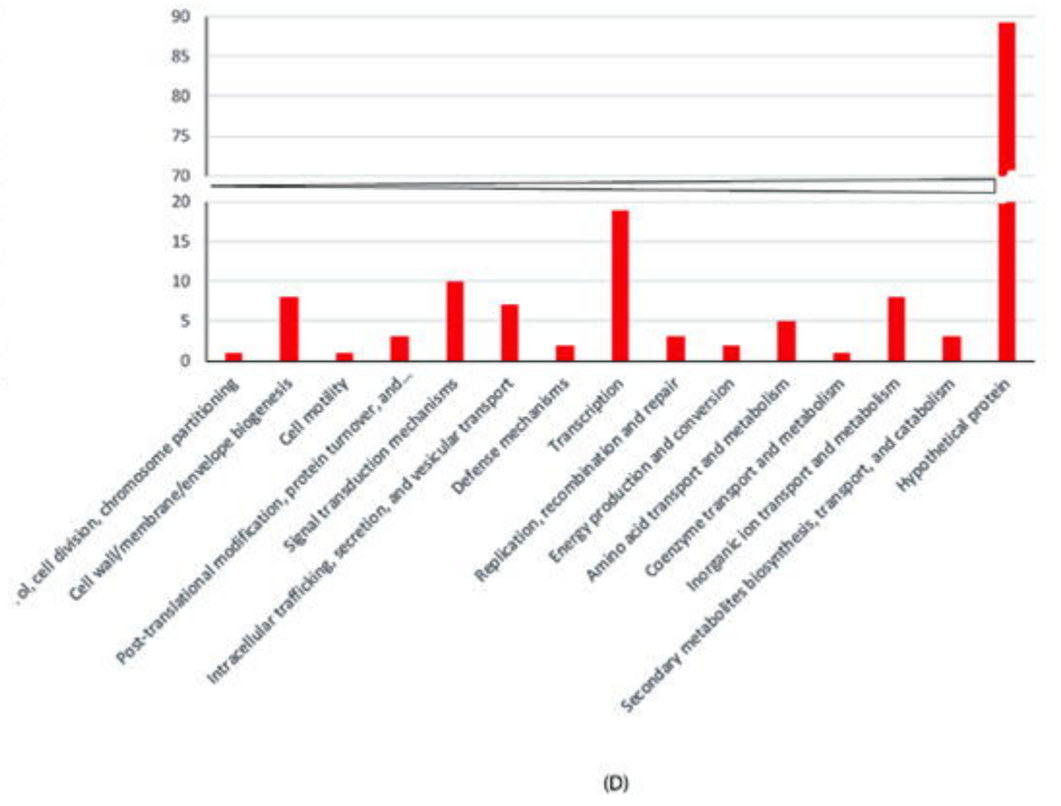
(A)


25
20
15
10
5
0
Cell wall/membrane/envelope biogenesis
Cell motility
Post-translational modification, protein turnover, and chaperones
Signal transduction mechanisms
Translation, ribosomal structure and biogenesis
Energy production and conversion
Amino acid transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis, transport, and catabolism
Hypothetical protein

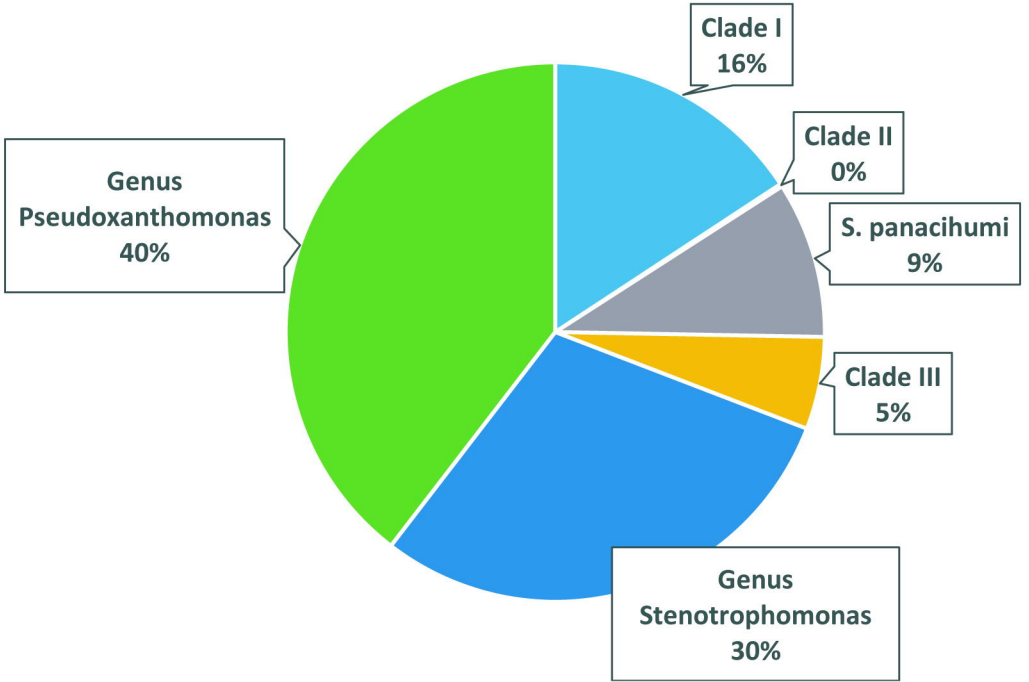
(B)

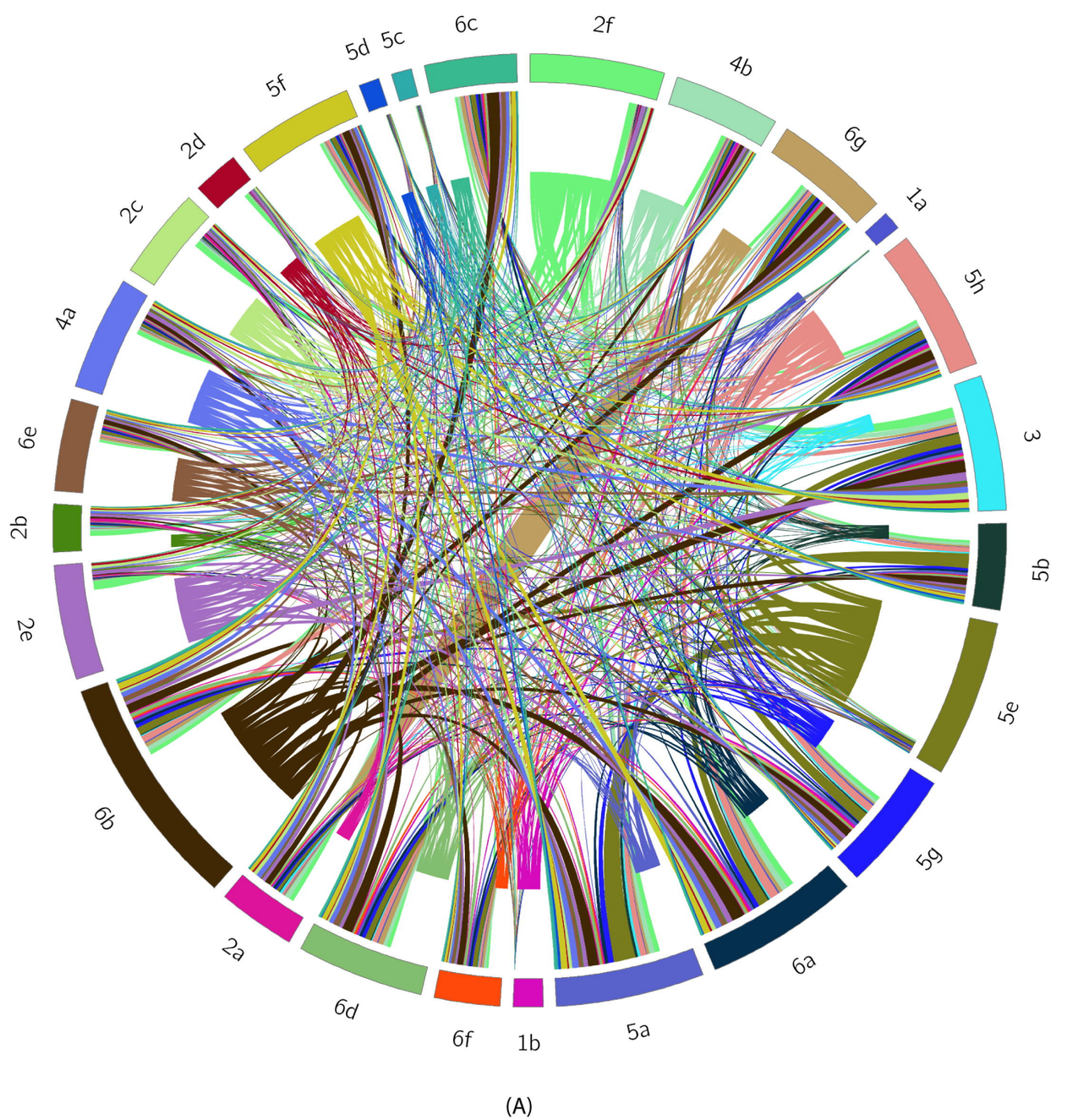

120
115
110
105
100
20
15
10
5
0
Cell cycle control, cell division, chromosome partitioning
Cell wall/membrane/envelope biogenesis
Post-translational modification, protein turnover, and chaperones
Signal transduction mechanisms
Intracellular trafficking, secretion, and vesicular transport
Defense mechanisms
Transcription
Replication, recombination and repair
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis, transport, and catabolism
Hypothetical protein

(C)

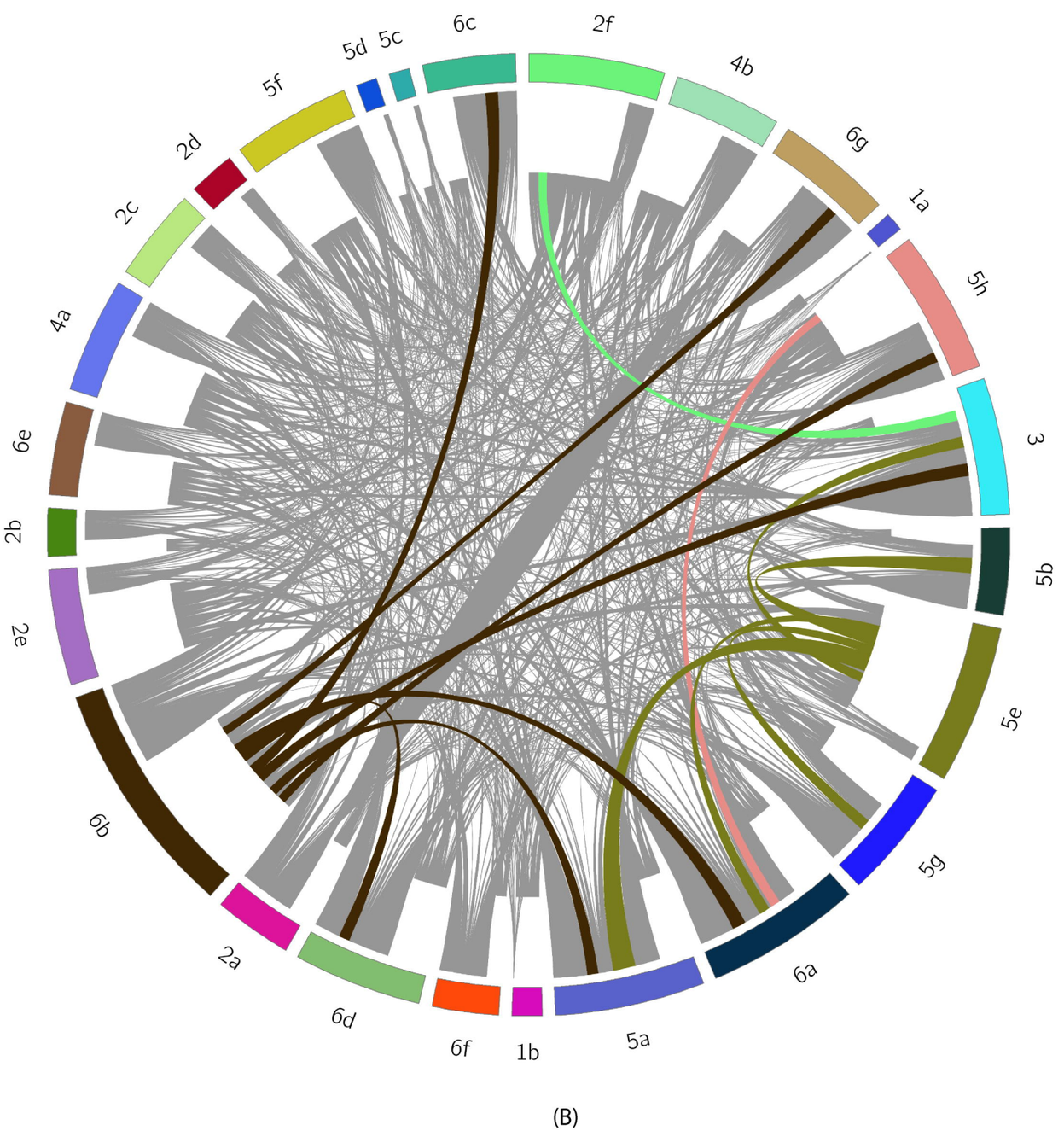

90
85
80
75
70
20
15
10
5
0
, cell division, chromosome partitioning
Cell wall/membrane/envelope biogenesis
Cell motility
Post-translational modification, protein turnover, and...
Signal transduction mechanisms
Intracellular trafficking, secretion, and vesicular transport
Defense mechanisms
Transcription
Replication, recombination and repair
Energy production and conversion
Amino acid transport and metabolism
Coenzyme transport and metabolism
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis, transport, and catabolism
Hypothetical protein

(D)

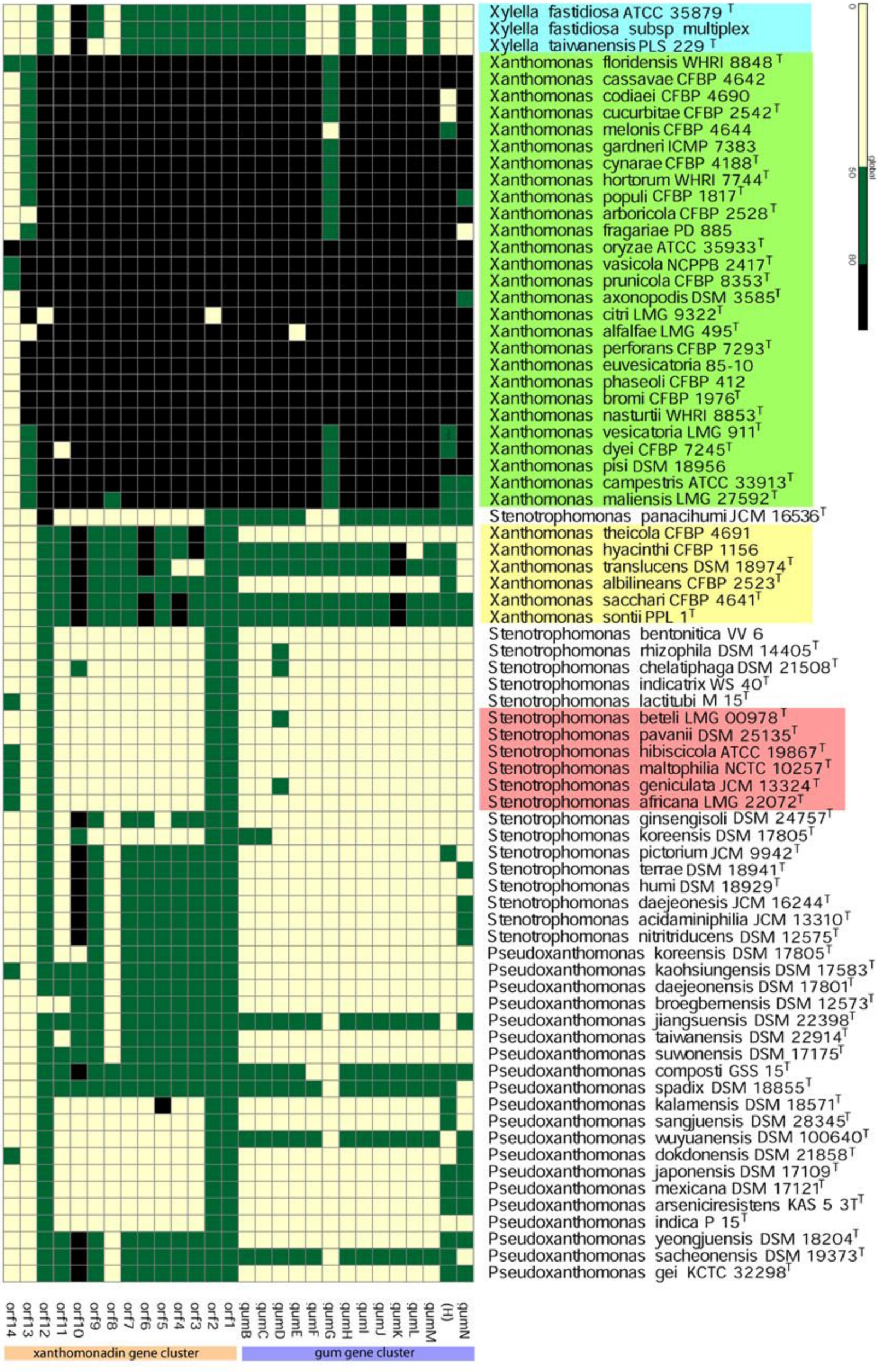

Clade I
16%
Clade II
0%
S. panacihumi
9%
Clade III
5%
Genus
Stenotrophomonas
30%
Genus
Pseudoxanthomonas
40%


6c
2f
4b
6g
1a
5h
3
5b
5e
5g
6a
5a
1b
6f
6d
2a
6b
2e
2b
6e
4a
2c
2d
5f
5d
5c

(A)


6c
2f
4b
6g
1a
5h
3
5b
5e
5g
6a
5a
1b
6f
6d
2a
6b
2e
2b
6e
4a
2c
2d
5f
5d
5c

(B)


Xylella fastidiosa ATCC 35879 T
Xylella fastidiosa subsp multiplex
Xylella taiwanensis PLS 229 T
Xanthomonas floridensis WHRI 8848 T
Xanthomonas cassavae CFBP 4642
Xanthomonas codiaei CFBP 4690
Xanthomonas cucurbitae CFBP 2542 T
Xanthomonas melonis CFBP 4644
Xanthomonas gardneri ICMP 7383
Xanthomonas cynarae CFBP 4188 T
Xanthomonas hortorum WHRI 7744 T
Xanthomonas populi CFBP 1817 T
Xanthomonas arboricola CFBP 2528 T
Xanthomonas fragariae PD 885
Xanthomonas oryzae ATCC 35933 T
Xanthomonas vasicola NCPPB 2417 T
Xanthomonas prunicola CFBP 8353 T
Xanthomonas axonopodis DSM 3585 T
Xanthomonas citri LMG 9322 T
Xanthomonas alfalfae LMG 495 T
Xanthomonas perforans CFBP 7293 T
Xanthomonas euvesicatoria 85-10
Xanthomonas phaseoli CFBP 412
Xanthomonas bromi CFBP 1976 T
Xanthomonas nasturtii WHRI 8853 T
Xanthomonas vesicatoria LMG 911 T
Xanthomonas dyei CFBP 7245 T
Xanthomonas pisi DSM 18956
Xanthomonas campestris ATCC 33913 T
Xanthomonas maliensis LMG 27592 T
Stenotrophomonas panacihumi JCM 16536 T
Xanthomonas theicola CFBP 4691
Xanthomonas hyacinthi CFBP 1156
Xanthomonas translucens DSM 18974 T
Xanthomonas albilineans CFBP 2523 T
Xanthomonas sacchari CFBP 4641 T
Xanthomonas sontii PPL 1 T
Stenotrophomonas bentonitica VV 6
Stenotrophomonas rhizophila DSM 14405 T
Stenotrophomonas chelatiphaga DSM 21508 T
Stenotrophomonas indicatrix WS 40 T
Stenotrophomonas lactitubi M 15 T
Stenotrophomonas beteli LMG 00978 T
Stenotrophomonas pavanii DSM 25135 T
Stenotrophomonas hibiscicola ATCC 19867 T
Stenotrophomonas maltophilia NCTC 10257 T
Stenotrophomonas geniculata JCM 13324 T
Stenotrophomonas africana LMG 22072 T
Stenotrophomonas ginsengisoli DSM 24757 T
Stenotrophomonas koreensis DSM 17805 T
Stenotrophomonas pictorium JCM 9942 T
Stenotrophomonas terrae DSM 18941 T
Stenotrophomonas humi DSM 18929 T
Stenotrophomonas daejeonesis JCM 16244 T
Stenotrophomonas acidaminiphilia JCM 13310 T
Stenotrophomonas nitritriducens DSM 12575 T
Pseudoxanthomonas koreensis DSM 17805 T
Pseudoxanthomonas kaohsiungensis DSM 17583 T
Pseudoxanthomonas daejeonensis DSM 17801 T
Pseudoxanthomonas broegbernensis DSM 12573 T
Pseudoxanthomonas jiangsuensis DSM 22398 T
Pseudoxanthomonas taiwanensis DSM 22914 T
Pseudoxanthomonas suwonensis DSM 17175 T
Pseudoxanthomonas composti GSS 15 T
Pseudoxanthomonas spadix DSM 18855 T
Pseudoxanthomonas kalamensis DSM 18571 T
Pseudoxanthomonas sangjuensis DSM 28345 T
Pseudoxanthomonas wuyuanensis DSM 100640 T
Pseudoxanthomonas dokdonensis DSM 21858 T
Pseudoxanthomonas japonensis DSM 17109 T
Pseudoxanthomonas mexicana DSM 17121 T
Pseudoxanthomonas arseniciresistens KAS 5 3T T
Pseudoxanthomonas indica P 15 T
Pseudoxanthomonas yeongjuensis DSM 18204 T
Pseudoxanthomonas sacheonensis DSM 19373 T
Pseudoxanthomonas gei KCTC 32298 T
0
50
80
qkbtot
orf14
orf13
orf12
orf11
orf10
orf9
orf8
orf7
orf6
orf5
orf4
orf3
orf2
orf1
gumB
gumC
gumD
gumE
gumF
gumG
gumH
gumI
gumJ
gumK
gumL
gumM
gumN (H)
xanthomonadin gene cluster
gum gene cluster