**Standards in
Genomic Sciences**

SHORT GENOME REPORT

Open Access

CrossMark

Genome sequence of the model plant pathogen *Pectobacterium carotovorum* SCC1

Outi Niemi[1,2]* , Pia Laine[3], Patrik Koskinen[1,3], Miia Pasanen[4], Ville Pennanen[1,2], Heidi Harjunpää[1,2], Johanna Nykyri[4], Liisa Holm[1,3], Lars Paulin[3], Petri Auvinen[3], E. Tapio Palva[1,2] and Minna Pirhonen[4]

Abstract

Bacteria of the genus *Pectobacterium* are economically important plant pathogens that cause soft rot disease on a wide variety of plant species. Here, we report the genome sequence of *Pectobacterium carotovorum* strain SCC1, a Finnish soft rot model strain isolated from a diseased potato tuber in the early 1980's. The genome of strain SCC1 consists of one circular chromosome of 4,974,798 bp and one circular plasmid of 5524 bp. In total 4451 genes were predicted, of which 4349 are protein coding and 102 are RNA genes.

Keywords: *Pectobacterium*, Soft rot, Plant pathogen, Necrotroph, Potato, Finland

Introduction

Pectobacterium species are economically important plant pathogens that cause soft rot and blackleg disease on a range of plant species across the world [1, 2]. The main virulence mechanism employed by *Pectobacterium* is the secretion of vast amounts of plant cell wall-degrading enzymes [1, 3]. Due to their ability to effectively macerate plant tissue for acquisition of nutrients, *Pectobacterium* species are considered classical examples of necrotrophic plant pathogens. Among the *Pectobacterium* species, *P. carotovorum* has the widest host range while potato is the most important crop affected in temperate regions [1, 4]. *P. carotovorum* strain SCC1 was isolated from a diseased potato tuber in Finland in the early 1980's [5]. It is highly virulent on model plant hosts such as tobacco (*Nicotiana tabacum*) and thale cress (*Arabidopsis thaliana*) as well as on the original host, potato (*Solanum tuberosum*). For three decades, the strain has been used as a model strain in the study of virulence mechanisms of *Pectobacterium* as well as in the study of plant defense mechanisms against necrotrophic

plant pathogens ([e.g. [6–13]). Here we describe the annotated genome sequence of *P. carotovorum* strain SCC1.

Organism information
Classification and features

P. carotovorum strain SCC1 is a Gram-negative, motile, non-sporulating, and facultatively anaerobic bacterium that belongs to the order of *Enterobacterales* within the class of *Gammaproteobacteria*. Cells of strain SCC1 are rod shaped with length of approximately 2 μm in the exponential growth phase (Fig. 1). Strain SCC1 is pathogenic causing soft rot disease in plants. It was originally isolated from a diseased potato tuber in Finland in 1982 [5]. It also provokes maceration symptoms on model plants Arabidopsis, tobacco, and tomato (*Solanum lycopersicum*), and is used as a soft rot model in research.
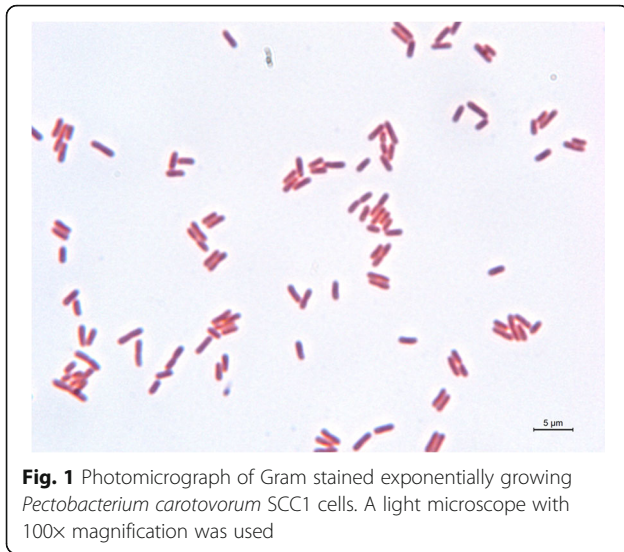
Strain SCC1 has previously been described belonging to *P. carotovorum* subsp. *carotovorum* based on biochemical properties such as its ability to grow at +37 °C and in 5% NaCl, its sensitivity to erythromycin, its ability to assimilate lactose, melibiose and raffinose but not sorbitol, and its inability to produce reducing sugars from sucrose and acid from α-methyl glucoside [14]. A phylogenetic tree generated based on seven housekeeping genes (*dnaN, fusA, gyrB, recA, rplB, rpoS* and *gyrA*) clusters strain SCC1 together with other *P. carotovorum* strains (Fig. 2). However, sequence based

* Correspondence: outi.niemi@helsinki.fi
[1]Division of Genetics, Department of Biosciences, University of Helsinki, Helsinki, Finland
[2]Viikki Plant Science Centre, University of Helsinki, Helsinki, Finland
Full list of author information is available at the end of the article

Niemi et al. Standards in Genomic Sciences (2017) 12:87

Page 2 of 8

Fig. 1 Photomicrograph of Gram stained exponentially growing *Pectobacterium carotovorum* SCC1 cells. A light microscope with 100× magnification was used

phylogenetic analysis was inconclusive regarding the subspecies status. Overall, the phylogeny of *Pectobacterium* species and subspecies is currently in turmoil and assigning strains to subspecies is challenging [15].

P. carotovorum strain SCC1 has been deposited at the International Center for Microbial Resources - French collection of plant-associated bacteria (accession: CFBP 8537). MIGS of strain SCC1 is summarized in Table 1.

## Genome sequencing information
Genome project history
P. carotovorum strain SCC1 has been used as a model soft rot pathogen in the field of plant-pathogen interactions ever since its isolation in the 1980's. The

sequencing of the genome of strain SCC1 was initiated in 2008 in order to further facilitate its use as a model pathogen.

The project was carried out jointly by the Institute of Biotechnology, Department of Biosciences and Department of Agricultural Sciences at the University of Helsinki, Finland. The genome was sequenced, assembled and annotated. The final sequence contains two scaffolds representing one chromosome and one plasmid. The sequence of the chromosome contains one gap of estimated length of 3788 bp. The genome sequence is deposited in GenBank under the accession numbers CP021894 (chromosome) and CP021895 (plasmid). Summary information of the project is presented in Table 2.

Growth conditions and genomic DNA preparation
After isolation from potato in 1982, *P. carotovorum* strain SCC1 has been stored in 22% glycerol at −80 °C. For preparation of genomic DNA, the strain was first grown overnight on solid LB medium (10 g tryptone, 5 g yeast extract, 10 g NaCl, and 15 g agar per one liter of medium) at 28 °C. A single colony was then picked and grown overnight in 10 ml of liquid LB medium at 28 °C with shaking. Cells were harvested by centrifugation for 20 min at 3200 g at 4 °C and resuspended into TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA). SDS (5% w/v) and Proteinase K (1 mg/ml) were used to break the cells for one hour at 50 °C. Genomic DNA was extracted using phenol-chloroform purification followed by ethanol precipitation. The quantity and quality of the DNA was assessed by spectrophotometry and agarose gel electrophoresis.
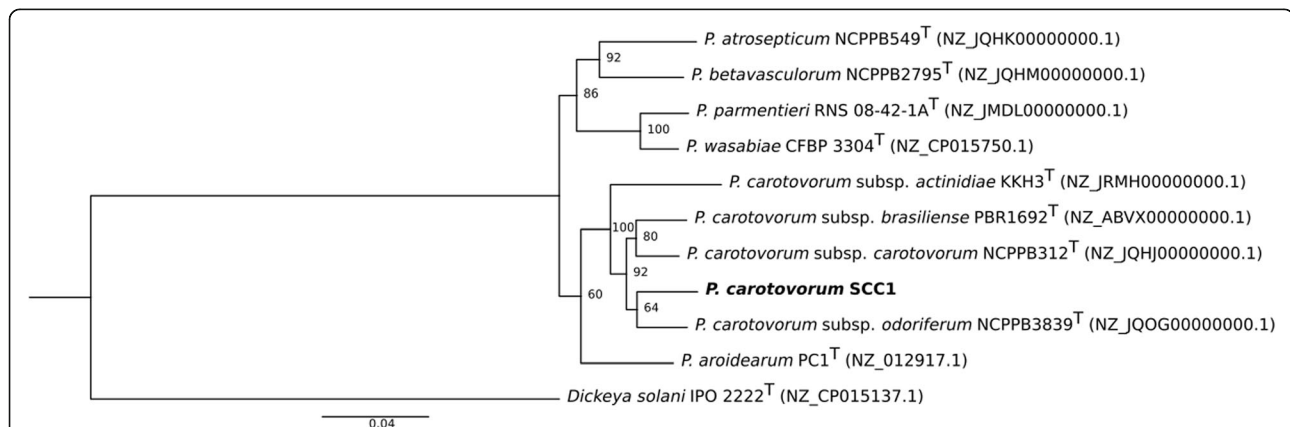
Fig. 2 Maximum likelihood tree of *Pectobacterium carotovorum* SCC1 and other closely related *Pectobacterium* strains. The phylogenetic tree was constructed from the seven housekeeping genes (*dnaN, fusA, gyrB, recA, rplB, rpoS* and *gyrA*). The concatenated sequences were aligned using MAFFT multiple sequence alignment program (version 7) with default parameters [42]. The phylogenetic tree was built in RAxML (Randomized Axelerated Maximum Likelihood) program with Maximum likelihood (ML) inference [43]. 88 different nucleotide substitution models were tested with jModelTest 2.0 and the best model was selected using Akaike information criterion (AIC) [44]. Bootstrap values from 1000 replicates are shown in each branch. *Dickeya solani* IPO2222 was used as the outgroup. Type strains are marked with T after the strain name. GenBank accession numbers are presented in the parentheses. The scale bar indicates 0.04 substitutions per nucleotide position

Niemi *et al. Standards in Genomic Sciences* (2017) 12:87

Page 3 of 8

Table 1 Classification and general features of *Pectobacterium carotovorum* strain SCC1 [46]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain *Bacteria*	TAS [47]
		Phylum *Proteobacteria*	TAS [48]
		Class *Gammaproteobacteria*	TAS [49, 50]
		Order *Enterobacterales*	TAS [51]
		Family *Pectobacteriaceae*	TAS [51]
		Genus *Pectobacterium*	TAS [52, 53]
		Species *Pectobacterium carotovorum*	TAS [52, 54]
		Strain: SCC1 (CFBP 8537)	TAS [5]
	Gram stain	Negative	IDA
	Cell shape	Rod	IDA
	Motility	Motile	IDA
	Sporulation	Non-sporulating	NAS [51]
	Temperature range	Mesophile, able to grow at 37 °C	TAS [14]
	Optimum temperature	~28 °C	IDA
	pH range; Optimum	Unknown	
	Carbon source	Sucrose, lactose, melibiose, raffinose	IDA,TAS [14]
MIGS-6	Habitat	Potato	TAS [5]
MIGS-6.3	Salinity	Able to grow in 5% NaCl	TAS [14]
MIGS-22	Oxygen requirement	Facultatively anaerobic	NAS [51]
MIGS-15	Biotic relationship	Free-living	NAS
MIGS-14	Pathogenicity	Pathogenic	NAS [53]
MIGS-4	Geographic location	Finland	TAS [5]
MIGS-5	Sample collection	1982	NAS
MIGS-4.1	Latitude	60° 13′ 36.15″ N	NAS
MIGS-4.2	Longitude	25° 00′ 54.77″ E	NAS
MIGS-4.4	Altitude	Unknown	

[a]Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [55]

Table 2 Project information

MIGS ID	Property	Term
MIGS 31	Finishing quality	One gap remaining, otherwise finished
MIGS-28	Libraries used	Standard 454 and Solid libraries
MIGS 29	Sequencing platforms	454, SOLiD, Sanger
MIGS 31.2	Fold coverage	Chromosome 40×, plasmid 67×
MIGS 30	Assemblers	gsAssembler v 1.1.03.24
MIGS 32	Gene calling method	Prodigal
	Locus Tag	SCC1
	Genbank ID	CP021894, CP021895
	GenBank Date of Release	July 27, 2017
	GOLD ID	
	BIOPROJECT	PRJNA379819
MIGS 13	Source Material Identifier	CFBP 8537
	Project relevance	Plant pathogen

Niemi *et al. Standards in Genomic Sciences* (2017) 12:87

Page 4 of 8

Genome sequencing and assembly

Genome sequencing was performed at DNA and Genomics Laboratory, Institute of Biotechnology, University of Helsinki, Finland. Genomic DNA was sequenced using 454 (454 Life Sciences/Roche), SOLiD3 (Life Technologies) and ABI 3130xl Genetic Analyzer (Life Technologies) instruments. DNA was fragmented into approximate size of 800 bp using Nebulizer (Roche) followed by standard fragment 454 library with the GS FLX series reagents. For the SOLiD library genomic DNA was fragmented with a Covaris S2 Sonicator (Covaris Inc.) to approximate size of 250 bp. The library was prepared using the SOLiD library kit (Life technologies).Newbler (version 1.1) was used to assemble 366,453 pyrosequencing reads (77,6 Mbp) in approximate length of 240 bp with default settings into 100 large (>1000 bp) contigs. GAP4 program (Staden package) was used for contig editing, primer design for PCRs and primer walking, and finishing the genome. Gaps were closed using PCR and traditional primer walking Sanger sequencing method. Finally, SOLiD reads were mapped to the genome and fifteen single genomic positions were fixed. Final sequencing coverages were 40× in genome and 67× in plasmid sequences.

Genome annotation

Coding sequences were predicted using the Prodigal gene prediction tool [16]. GenePRIMP [17] was run to correct systematic errors made by Prodigal and to reanalyze the remaining intergenic regions for missed CDSs. Functional annotation for the predicted genes was performed using the PANNZER annotation tool [18]. The annotation was manually curated with information from publications and the following databases: COG [19], KEGG [20], CDD [21], UniProt and NCBI nonredundant protein sequences. To identify RNA genes, RNAmmer v1.2 [22] (rRNAs) and tRNAscan-SE [23] (tRNAs) were used. Clusters of Orthologous Groups assignments and Pfam domain predictions were done using the WebMGA server [24]. Transmembrane helices were predicted with TMHMM [25] and Phobius [26]. For signal peptide prediction, SignalP 4.1 [27] was used. CRISPRFinder [28] was used to detect Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs).

Genome properties

The genome of *P. carotovorum* SCC1 consists of one circular 4,974,798 bp chromosome and one circular 5524 bp plasmid (Table 3, Fig. 3). The total genome size is 4,980,322 bp with an overall G + C content of 51.85% (Table 4). A total of 4451 genes were predicted, out of which 4440 are chromosomal and 11 reside on the plasmid. 4349 (97.71%) genes are protein coding and 102 (2.29%) are RNA genes (77 tRNA, 22 rRNA, and 3 other

Table 3 Summary of *P. carotovorum* SCC1 genome: one chromosome and one plasmid

Label	Size (Mb)	Topology	INSDC identifier	RefSeq ID
Chromosome	4.974798	Circular	CP021894	
Plasmid pSCC1	0.005524	Circular	CP021895	

RNA genes). Of the 4349 protein coding genes, 3812 (87.65%) could be assigned to COG functional categories (Table 5).
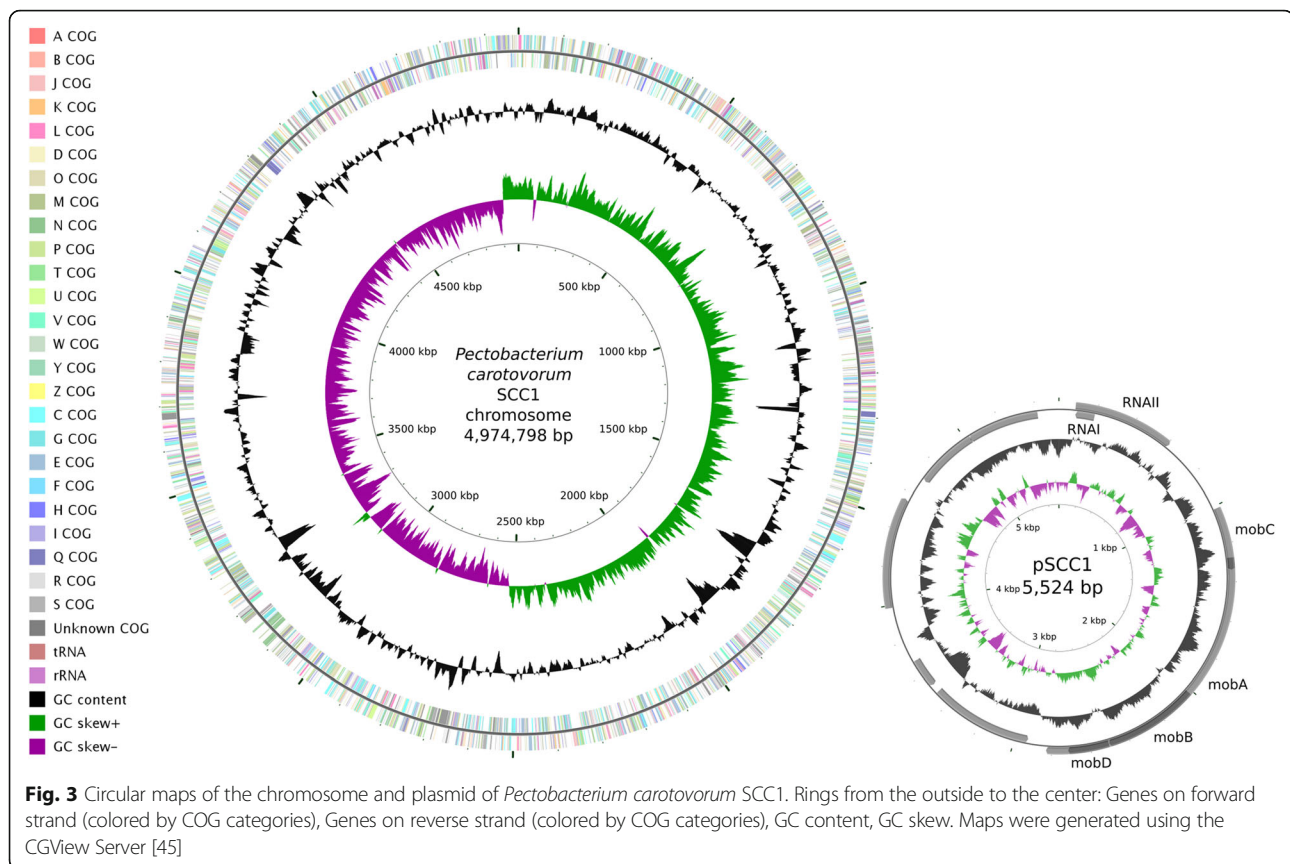
Insights from the genome sequence

P. carotovorum strain SCC1 harbors a small cryptic plasmid of 5524 bp, pSCC1. The plasmid contains sequences for RNAI and RNAII, two non-coding RNAs involved in replication initiation and control in enterobacterial RNA priming plasmids such as ColE1 [29]. A similar replication region has previously been described in the small cryptic plasmid pEC3 of *P. carotovorum* subsp. *carotovorum* strain IFO3380 [30]. In addition to the two RNA genes, pSCC1 was predicted to contain nine protein-coding genes. Four of these (*mobABCD*) encode mobilization proteins. The *mob* locus is required for mobilization of non-self-transmissible plasmids and is found on many enterobacterial plasmids including pEC3 [31]. No function could be assigned to the remaining five genes on pSCC1. One of them, SCC1_4463, is very similar to genes found in many *Enterobacteriaceae* genomes, especially those of genera *Enterobacter*, *Escherichia* and *Salmonella*, whereas similar genes to the other four on pSCC1 are not widely present in other sequenced genomes.

Pectobacterium infection is characterized by maceration symptoms caused by the secretion of a large arsenal of plant cell wall-degrading enzymes. Accordingly, the genome of *P. carotovorum* strain SCC1 was found to contain genes for eleven pectate lyases (*pelABCILWXZ*, *hrpW*, SCC1_1311, and SCC1_2381), one pectin lyase (*pnl*), four polygalacturonases (*pehAKNX*), one oligogalacturonate lyase (*ogl*), three cellulases (*celSV*, *bcsZ*), one rhamnogalacturonate lyase (*rhiE*), two pectin methylesterases (*pemAB*), and two pectin acetylesterases (*paeXY*). In addition, the genome harbors two genes encoding proteases previously characterized as plant cell wall-degrading enzymes (*prt1*, *prtW*) as well as a number of putative proteases, some of which may function in plant cell wall degradation. Different *Pectobacterium* species and strains have been found to harbor very similar collections of plant cell wall-degrading enzymes [32], and the number and types of enzymes in the genome of strain SCC1 fit this picture well.

Protein secretion plays an essential role in soft rot pathogenesis [33]. The most important secretion system in *Pectobacterium* is the type II secretion system, also

Niemi *et al. Standards in Genomic Sciences* (2017) 12:87

Page 5 of 8

Fig. 3 Circular maps of the chromosome and plasmid of *Pectobacterium carotovorum* SCC1. Rings from the outside to the center: Genes on forward strand (colored by COG categories), Genes on reverse strand (colored by COG categories), GC content, GC skew. Maps were generated using the CGView Server [45]

known as the Out system (*outCDEFGHIJKLMN*), which transports proteins from the periplasmic space into the extracellular environment [34]. It is responsible for the secretion of most plant cell wall-degrading enzymes such as pectinases and cellulases as well as some other

Table 4 Genome statistics

Attribute	Value	% of Total
Genome size (bp)	4,980,322	100.00
DNA coding (bp)	4,314,063	86.62
DNA G + C (bp)	2,580,564	51.85
DNA scaffolds	2	
Total genes	4451	100.00
Protein coding genes	4349	97.71
RNA genes	102	2.29
Pseudo genes	NA	NA
Genes in internal clusters	NA	NA
Genes with function prediction	3955	88.86
Genes assigned to COGs	3812	85.64
Genes with Pfam domains	3782	84.97
Genes with signal peptides	428	9.62
Genes with transmembrane helices	939	21.10
CRISPR repeats	5	

virulence factors such as the necrosis-inducing protein Nip [33, 35]. Furthermore, *Pectobacterium* genomes typically harbor multiple type I secretion systems, which secrete proteases and adhesins [33]. At least four type I secretion systems are encoded in the genome of *P. carotovorum* SCC1 (*prtDEF*, SCC1_1144–1146, SCC1_1589–1591, and SCC1_3286–3288). Strain SCC1 also harbors a type III secretion system cluster (SCC1_2406–2432), which has previously been characterized in this strain and shown to affect the speed of symptom development during infection [6, 36]. Overall, the role of the type III secretion system in *Pectobacterium* is not well understood and *P. wasabiae* and *P. parmentieri* seem to lack it completely [32, 37]. The type IV secretion system has been shown to have a minor contribution to virulence of *P. atrosepticum* [38]. However, it is sporadically distributed among *Pectobacterium* strains [33], and no type IV secretion genes could be found from the genome of *P. carotovorum* SCC1. Finally, the type VI secretion system has also been shown to have a small effect on virulence at least in some *Pectobacterium* species [32, 39]. In *P. carotovorum* SCC1, one type VI secretion system cluster is present in the genome (SCC1_0988–1002).

Soft rot pathogens have been suggested to be able to use insect vectors in transmission, and indeed, certain *P.*

Niemi *et al. Standards in Genomic Sciences* (2017) 12:87

Page 6 of 8

Table 5 Number of genes associated with general COG functional categories

Code	Value	%age	Description
J	183	4.21	Translation, ribosomal structure and biogenesis
A	2	0.05	RNA processing and modification
K	332	7.63	Transcription
L	161	3.70	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	40	0.92	Cell cycle control, Cell division, chromosome partitioning
V	61	1.40	Defense mechanisms
T	242	5.56	Signal transduction mechanisms
M	241	5.54	Cell wall/membrane biogenesis
N	114	2.62	Cell motility
U	122	2.81	Intracellular trafficking and secretion
O	152	3.50	Posttranslational modification, protein turnover, chaperones
C	248	5.70	Energy production and conversion
G	376	8.65	Carbohydrate transport and metabolism
E	435	10.00	Amino acid transport and metabolism
F	94	2.16	Nucleotide transport and metabolism
H	177	4.07	Coenzyme transport and metabolism
I	103	2.37	Lipid transport and metabolism
P	318	7.31	Inorganic ion transport and metabolism
Q	67	1.54	Secondary metabolites biosynthesis, transport and catabolism
R	445	10.23	General function prediction only
S	358	8.23	Function unknown
–	537	12.35	Not in COGs

The total is based on the total number of protein coding genes in the genome

carotovorum strains can infect *Drosophila* flies and persist in their guts [40]. This ability has been linked to the Evf (*Erwinia* virulence factor) protein [41]. The *evf* gene is present in the genome of *P. carotovorum* SCC1 suggesting that the strain may have the ability to interact with insects.

Conclusions

In this study, we presented the annotated genome sequence of the pectinolytic plant pathogen *Pectobacterium carotovorum* SCC1 consisting of a chromosome of 4,974,798 bp and a small cryptic plasmid of 5524 bp. Strain SCC1 was originally isolated from a diseased potato tuber and it has been used as a model strain to study interactions between soft rot pathogens and their host plants for decades. In accordance with the pathogenic lifestyle, the genome of strain SCC1 was found to harbor a large arsenal of plant cell wall-degrading enzymes similar to other sequenced *Pectobacterium* genomes. In addition, an insect interaction gene, *evf*, is present in the genome of strain SCC1 suggesting the possibility of insects as vectors or alternative hosts for this strain. The genome sequence will drive further the use of *P. carotovorum* SCC1 as a model plant pathogen.

Authors' contributions
ETP and MPi initiated the study and provided the strain and background information. HH isolated the genomic DNA. PL, PA and LP designed sequencing strategy and performed genome sequencing and assembly. PK, LH and ON annotated the genome. ON manually corrected the functional annotation. MPa and MPi conducted the phylogenetic analysis. ON, MPa and MPi performed biological experiments. ON, VP, JN and MPi analysed the contents of the genome. ON wrote the manuscript. MPi, PL and MPa contributed to writing. All authors read and approved the final manuscript.

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.

Niemi *et al. Standards in Genomic Sciences* (2017) 12:87

Page 7 of 8

Author details
[1]Division of Genetics, Department of Biosciences, University of Helsinki, Helsinki, Finland. [2]Viikki Plant Science Centre, University of Helsinki, Helsinki, Finland. [3]Institute of Biotechnology, University of Helsinki, Helsinki, Finland. [4]Plant Pathology, Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland.

References
1. Pérombelon MCM. Potato diseases caused by soft rot erwinias: an overview of pathogenesis. Plant Pathol. 2002;51:1–12.
2. Czajkowski R, Pérombelon MCM, van Veen JA, van der Wolf JM. Control of blackleg and tuber soft rot of potato caused by *Pectobacterium* and *Dickeya* species: a review. Plant Pathol. 2011;60:999–1013.
3. Davidsson PR, Kariola T, Niemi O, Palva ET. Pathogenicity of and plant immunity to soft rot pectobacteria. Front Plant Sci. 2013;4:191.
4. Toth IK, Bell KS, Holeva MC, Birch PRJ. Soft rot erwiniae: from genes to genomes. Mol Plant Pathol. 2003;4:17–30.
5. Saarilahti HT, Palva ET. Major outer membrane proteins in the phytopathogenic bacteria *Erwinia carotovora* subsp. *carotovora* and subsp. *atroseptica*. FEMS Microbiol. Lett. 1986;35:267–70.
6. Rantakari A, Virtaharju O, Vähämiko S, Taira S, Palva ET, Saarilahti HT, et al. Type III secretion contributes to the pathogenesis of the soft-rot pathogen *Erwinia carotovora*: partial characterization of the hrp gene cluster. Mol PlantMicrobe Interact. 2001;14:962–8.
7. Li J, Brader G, Palva ET. The WRKY70 transcription factor: a node of convergence for Jasmonate-mediated and Salicylate-mediated signals in plant defense. Plant Cell. 2004;16:319–31.
8. Brader G, Sjöblom S, Hyytiäinen H, Sims-Huopaniemi K, Palva ET. Altering substrate chain length specificity of an acylhomoserine lactone synthase in bacterial communication. J Biol Chem. 2005;280:10403–9.
9. Kariola T, Brader G, Li J, Palva ET. Chlorophyllase 1, a damage control enzyme, affects the balance between defense pathways in plants. Plant Cell. 2005;17:282–94.
10. Kariola T, Brader G, Helenius E, Li J, Heino P, Palva ET. EARLY RESPONSIVE TO DEHYDRATION 15, a negative regulator of Abscisic acid responses in Arabidopsis. Plant Physiol. 2006;142:1559–73.
11. Kwon SJ, Jin HC, Lee S, Nam MH, Chung JH, Kwon SI, et al. GDSL lipase-like 1 regulates systemic resistance associated with ethylene signaling in Arabidopsis. Plant J. 2009;58:235–45.
12. Po-Wen C, Singh P, Zimmerli L. Priming of the Arabidopsis pattern-triggered immunity response upon infection by necrotrophic *Pectobacterium carotovorum* bacteria. Mol Plant Pathol. 2013;14:58–70.
13. Singh P, Kuo Y-C, Mishra S, Tsai C-H, Chien C-C, Chen C-W, et al. The Lectin receptor Kinase-VI.2 is required for priming and positively regulates Arabidopsis pattern-triggered immunity. Plant Cell. 2012;24:1256–70.
14. Pasanen M, Laurila J, Brader G, Palva ET, Ahola V, van der Wolf J, et al. Characterisation of *Pectobacterium wasabiae* and *Pectobacterium carotovorum subsp. carotovorum* isolates from diseased potato plants in Finland. Ann Appl Biol. 2013;163:403–19.
15. Zhang Y, Fan Q, Loria R. A re-evaluation of the taxonomy of phytopathogenic genera *Dickeya* and *Pectobacterium* using whole-genome sequencing data. Syst Appl Microbiol. 2016;39:252–9.
16. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
17. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, et al. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat Methods. 2010;7:455–7.
18. Koskinen P, Törönen P, Nokso-Koivisto J, Holm L. PANNZER: high-throughput functional annotation of uncharacterized proteins in an error-prone environment. Bioinformatics. 2015;31:1544–52.
19. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28:33–6.
20. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.
21. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res. 2011;39:D225–9.
22. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100–8.
23. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.
24. Wu S, Zhu Z, Fu L, Niu B, Li W. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics. 2011;12:444.
25. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567–80.
26. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Res. 2007;35:W429–32.
27. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods. 2011;8:785–6.
28. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007;35:W52–7.
29. Cesareni G, Helmer-Citterich M, Castagnoli L. Control of ColE1 plasmid replication by antisense RNA. Trends Genet. 1991;7:230–5.
30. Nomura N, Murooka Y. Characterization and sequencing of the region required for replication of a non-selftransmissible plasmid pEC3 isolated from *Erwinia carotovora* subsp. *carotovora*. J Ferment Bioeng. 1994;78:250–4.
31. Nomura N, Yamashita M, Murooka Y. Genetic organization of a DNA-processing region required for mobilization of a non-self-transmissible plasmid, pEC3, isolated from *Erwinia carotovora* subsp. *carotovora*. Gene. 1996;170:57–62.
32. Nykyri J, Niemi O, Koskinen P, Nokso-Koivisto J, Pasanen M, Broberg M, et al. Revised phylogeny and novel horizontally acquired virulence determinants of the model soft rot Phytopathogen *Pectobacterium wasabiae* SCC3193. PLoS Pathog. 2012;8:e1003013.
33. Charkowski A, Blanco C, Condemine G, Expert D, Franza T, Hayes C, et al. The role of secretion systems and small molecules in soft-rot enterobacteriaceae pathogenicity. Annu Rev Phytopathol. 2012;50:425–49.
34. Johnson TL, Abendroth J, Hol WGJ, Sandkvist M. Type II secretion: from structure to function. FEMS Microbiol Lett. 2006;255:175–86.
35. Laasik E, Põllumaa L, Pasanen M, Mattinen L, Pirhonen M, Mäe A. Expression of nipP.W of *Pectobacterium wasabiae* is dependent on functional flgKL flagellar genes. Microbiology. 2014;160:179–86.
36. Lehtimäki S, Rantakari A, Routtu J, Tuikkala A, Li J, Virtaharju O, et al. Characterization of the hrp pathogenicity cluster of *Erwinia carotovora subsp. carotovora*: high basal level expression in a mutant is associated with reduced virulence. Mol Gen Genomics. 2003;270:263–72.
37. Kim H-S, Ma B, Perna NT, Charkowski AO. Phylogeny and virulence of naturally occurring type III secretion system-deficient *Pectobacterium* strains. Appl Environ Microbiol. 2009;75:4539–49.
38. Bell KS, Sebaihia M, Pritchard L, Holden MTG, Hyman LJ, Holeva MC, et al. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora subsp. atroseptica* and characterization of virulence factors. Proc Natl Acad Sci U S A. 2004;101:11105–10.
39. Liu H, Coulthurst SJ, Pritchard L, Hedley PE, Ravensdale M, Humphris S, et al. Quorum sensing coordinates brute force and stealth modes of infection in the plant pathogen *Pectobacterium atrosepticum*. PLoS Pathog. 2008;4: e1000093.
40. Basset A, Khush RS, Braun A, Gardan L, Boccard F, Hoffmann JA, et al. The phytopathogenic bacteria *Erwinia carotovora* infects drosophila and activates an immune response. Proc Natl Acad Sci U S A. 2000;97:3376–81.
41. Basset A, Tzou P, Lemaitre B, Boccard F. A single gene that promotes interaction of a phytopathogenic bacterium with its insect vector, Drosophila Melanogaster. EMBO Rep. 2003;4:205–9.
42. Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. In: Posada D, editor. Bioinformatics for DNA sequence analysis. New York: Humana Press; 2009. p. 39–64.
43. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
44. Posada D. ModelTest server: a web-based tool for the statistical selection of models of nucleotide substitution online. Nucleic Acids Res. 2006;34:W700–3.
45. Grant JR, Stothard P. The CGView server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 2008;36:W181–4.
46. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26:541–7.

Niemi *et al. Standards in Genomic Sciences* (2017) 12:87

Page 8 of 8

47. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87:4576–9.

48. Garrity GM, Bell JA, Lilburn T. Phylum XIV. *Proteobacteria* Phyl. Nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's manual of systematic bacteriology. Volume 2 (part B). 2nd ed. New York: Springer; 2005. p. 1.

49. Garrity GM, Bell JA, Lilburn T. Class III. *Gammaproteobacteria* class. Nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's manual of systematic bacteriology. Volume 2 (part B). 2nd ed. New York: Springer; 2005. p. 1.

50. Williams KP, Kelly DP. Proposal for a new class within the phylum *Proteobacteria*, *Acidithiobacillia* classis nov., with the type order *Acidithiobacillales*, and emended description of the class *Gammaproteobacteria*. Int J Syst Evol Microbiol. 2013;63:2901–6.

51. Adeolu M, Alnajar S, Naushad S, Gupta RS. Genome-based phylogeny and taxonomy of the '*Enterobacteriales*': proposal for *Enterobacterales* ord. Nov. divided into the families *Enterobacteriaceae*, *Erwiniaceae* fam. Nov., *Pectobacteriaceae* fam. Nov., *Yersiniaceae* fam. Nov., *Hafniaceae* fam. Nov., *Morganellaceae* fam. Nov., and *Budviciaceae* fam. Nov. Int J Syst Evol Microbiol. 2016;66:5575–99.

52. Waldee EL. Comparative studies of some peritrichous phytopathogenic bacteria. Iowa state Coll. J Sci. 1945;19:435–84.

53. Hauben L, Moore ER, Vauterin L, Steenackers M, Mergaert J, Verdonck L, et al. Phylogenetic position of phytopathogens within the *Enterobacteriaceae*. Syst Appl Microbiol. 1998;21:384–97.

54. Gardan L, Gouy C, Christen R, Samson R. Elevation of three subspecies of *Pectobacterium carotovorum* to species level: *Pectobacterium atrosepticum* sp. nov., *Pectobacterium betavasculorum* sp. nov. and *Pectobacterium wasabiae* sp. nov. Int J Syst Evol Microbiol. 2003;53:381–91.

55. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet. 2000;25:25–9.