Virology 507 (2017) 40–52

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Phylogenetic study of recombinant strains of *Potato virus Y*

Kelsie J. Green[a,c], Celeste J. Brown[b,c], Stewart M. Gray[d], Alexander V. Karasev[a,c,*]

[a] Department of PSES, University of Idaho, Moscow, ID, USA
[b] Department of Biological Sciences, University of Idaho, Moscow, ID, USA
[c] Bioinformatics and Computational Biology Program, University of Idaho, Moscow, ID, USA
[d] USDA-ARS and Section of Plant Pathology and Plant-Microbe Biology, School of Integrated Plant Science, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

Keywords:
Potato virus Y
Recombination
Evolution of recombinants

ABSTRACT

Potato virus Y (PVY) exists as a complex of strains, including a growing number of recombinants. Evolution of PVY proceeds through accumulation of mutations and more rapidly through recombination. Here, the role of recombination in PVY evolution and the origin of common PVY recombinants were studied through whole genome analysis of 119 newly sequenced PVY isolates largely from U.S. potato, and subsequent combined phylogenetic and recombination analyses with an additional 166 whole PVY genomes from the GenBank database. Two novel PVY^{C} recombinants were sequenced and identified, along with one novel $PVY^{N:O}$ recombinant. Sequence diversity in the parental sequences made it possible to trace the origins of all recombinant types of PVY, which also showed remarkable sequence diversity in most cases. The results suggested that the common recombinant PVY strains originated more than once, from different parental sequences.

1. Introduction

PVY is the type member of the genus *Potyvirus*, family *Potyviridae*, and has a single-stranded, positive-sense RNA genome ca. 9.7-kb nucleotides in length packaged in a flexible, filamentous structure (Adams et al., 2012). The PVY genome encodes a single polyprotein which is cleaved co- and post-translationally into ten mature proteins by three virus-specific proteases (Adams et al., 2012; Dougherty and Carrington, 1988). The PVY genome has a 3′ poly(A) tail, and its 5′-terminus is blocked by a covalently linked protein VPg (Riechmann et al., 1992). Recently, an additional ORF (named PIPO, or P3N-PIPO) was reported in a different reading frame and its product was found to interact with protein P3 and assist with movement of the virus *in planta* (Chung et al., 2008; Wei et al., 2010). In nature, PVY is transmitted mechanically, by aphids in a non-persistent manner, and also vegetatively through seed potato tubers (Kerlan, 2006).

PVY isolated from potato exists as a complex of strains that are distinguished based on hypersensitive resistance (HR) response towards three potato genes: *Ny*, *Nc*, and *Nz*. Isolates of PVY eliciting HR in the presence of the Ny_{tbr} belong to the PVY^{O} strain, those eliciting HR in the presence of Nc_{tbr} are classified as PVY^{C}, and those eliciting HR in the presence of Nz_{tbr} are classified as PVY^{Z} (Cockerham, 1970; de Bokx and Huttinga, 1981; Jones, 1990; Singh et al., 2008; Kerlan et al., 2011; Karasev and Gray, 2013a, 2013b; Quintero-Ferrer et al., 2014; Chikh-Ali et al., 2014, 2016c; Kehoe and Jones, 2016). Strains PVY^{N} (also called European N, or $PVY^{Eu\text{-}N}$) and PVY^{E} are unable to elicit HR in the presence of any of the three *N* genes, but $PVY^{Eu\text{-}N}$ isolates induce vein necrosis in tobacco, while PVY^{E} isolates induce only mosaic and vein clearing (Cockerham, 1970; Kerlan et al., 1999; Singh et al., 2008; Galvino-Costa et al., 2012a).

Genomes of PVY^{O}, $PVY^{Eu\text{-}N}$, and PVY^{C} strains are defined as non-recombinant and are found to serve as parents for many recombinant structures, with PVY^{O} and $PVY^{Eu\text{-}N}$ being the parents in the vast majority of PVY isolated from potato (Glais et al., 2002; Lorenzen et al., 2006a; Ogawa et al., 2008, 2012; Hu et al., 2009b). Considerable diversity is found in sequences of PVY isolates from the PVY^{O} strain, including a distinct sub-lineage called PVY^{O}-O5 (or PVY^{O5}), which can also be distinguished biologically (Karasev et al., 2011; Nie et al., 2012). There are sixteen recombinant PVY types reported to date, which include nine relatively common recombinants found in many geographical locations, namely $PVY^{N:O}$, $PVY^{N\text{-}Wi}$, PVY^{NTNa}, PVY^{NTNb}, PVY-NE11, PVY^{E}, and PVY-SYR-I, -II, and -III (Lorenzen et al., 2006a, 2008; Hu et al., 2009b; Chikh Ali et al., 2007, 2010; Schubert et al., 2007; Galvino-Costa et al., 2012a, 2012b; Karasev and Gray, 2013a), and seven rare recombinant types found and reported only once or twice, namely $PVY^{N\text{-}Wi}$-156var, $PVY^{N\text{-}Wi}$-261-4, PVY-SCRI-N, PVY-FrN, PVY-Nicola, PVY-T13, and PVY-nnp (e.g. Lorenzen et al., 2006a; Schubert et al., 2007; Lorenzen et al., 2008; Chikh Ali et al.,

* Corresponding author at: Department of PSES, University of Idaho, Moscow, ID 83844-2339, USA.
E-mail address: akarasev@uidaho.edu (A.V. Karasev).

http://dx.doi.org/10.1016/j.virol.2017.03.018
Received 4 March 2017; Received in revised form 29 March 2017; Accepted 30 March 2017
Available online 08 April 2017

2010; Galvino-Costa et al., 2012a; Ogawa et al., 2008, 2012; Karasev and Gray, 2013a). Only a few recombinants/recombinant types have been subjected to a thorough biological characterization on potato indicators carrying different *N* genes. From these studies, PVY^{NTNa} was found to belong to the PVY^{Z} strain, while PVY^{E} was found to have a sophisticated recombinant structure with PVY^{NTNa} and PVY-NE11 serving as parents (Kerlan et al., 2011; Galvino-Costa et al., 2012a; Quintero-Ferrer et al., 2014).

Initial computational analysis of PVY whole genomes suggested that the number of recombinant patterns reported for PVY isolates from potato are relatively limited, and the positions of the main recombinant junctions (RJs) are remarkably conserved, although it is as yet unclear why (Hu et al., 2009b). The driving forces for the emergence and survival of the PVY recombinants have not yet been elucidated, although there were no special sequences or RNA secondary structures found associated with the most common RJs (Hu et al., 2009b).

Evolution of PVY, and PVY recombinants in particular, was addressed in several recent attempts to re-create phylogenetic relationships between various virus recombinants (Ogawa et al., 2008, 2012; Karasev et al., 2011; Visser et al., 2012; Quenouille et al., 2013; Kehoe and Jones, 2016; Gibbs et al., 2017). Phylogenetic studies of PVY recombinants have been challenging because of the limited number of whole genomes available (Karasev et al., 2011; Ogawa et al., 2012; Visser et al., 2012; Kehoe and Jones, 2016), difficulties in accounting for recombination in building trees, ensuing necessities to analyze genome segments in order to infer the evolution of the entire genome (Karasev et al., 2011; Ogawa et al., 2012), and assumptions of monophyletic origins of PVY recombinants (Visser et al., 2012; Gibbs et al., 2017).

Here, we describe a large-scale sequencing project focused on a set of relatively well-characterized PVY strains isolated largely from North American field-grown potatoes. This characterization included serological typing with a set of monoclonal antibodies in TAS-ELISA, molecular typing by multiplex RT-PCR, and subsequent sequence analysis. One hundred nineteen whole genome sequences, together with 166 complete PVY sequences extracted from the GenBank database, were subjected to phylogenetic and recombination analysis. In this study, five common recombinant sections were individually analyzed for this collection of 285 whole-genome sequences of PVY isolates to determine their origins and evolutionary relationships.

2. Results

2.1. Strain identification and genome sequencing of PVY isolates

The 119 isolates of PVY collected from field grown potato were characterized by serological profiling and RT-PCR (see Materials and Methods) to classify them into strains or strain variants. Four isolates were identified as $PVY^{Eu\text{-}N}$ and fifty were identified as PVY^{O}, with 27 of these classified as PVY^{O}-O5. The remaining 65 isolates were identified as recombinant strains; 15 were PVY^{NTNa} isolates, eight were PVY-NE11, 22 were $PVY^{N:O}$, 17 were $PVY^{N\text{-}Wi}$ including one $PVY^{N\text{-}Wi}$-261-4-like isolate (Pondo4). Three isolates did not fit into any of the previously identified strain categories defined by serology and RT-PCR assays, and thus were designated as “unclassified” (Table 1).

The viral genomes were sequenced to provide further information about recombinant type. To facilitate sequencing, pairs of primers for amplifying recombinant genomes were designed to supplement previously published primer sets. Three primer groups were used for amplifying sequencing regions: a primer group designed for N-serotype strains (or “N-types”), e.g. $PVY^{Eu\text{-}N}$, $PVY^{NA\text{-}N}$, PVY-NE11, or sections of recombinants with these strains as parents; a primer group designed for O-serotype strains (or “O-types”), e.g. PVY^{O}, PVY^{O}-O5, or sections of recombinants with these strains as parents; and a collection of primers not intended to be strain-specific (Table 2). Most of the O- and N-type primers were found not to be strain-specific, however. The primer set was developed to produce ca. 1000-nt PCR fragments that overlapped each other for about 500-nt. Sequencing from both termini produced good overlap in the middle of the 1-kb PCR fragment. The entire set of all sequencing primers was used for each isolate, regardless of strain type. For most isolates, this was sufficient to amplify the entire 9.7-kb genome of PVY with a minimum of 4-fold sequence coverage. In the few cases where coverage was not adequate with this set alone, isolate-specific primer pairs were designed for each remaining sequence gap. The sequences were deposited in the GenBank and the corresponding accession numbers are listed in Table 1. In addition to the 119 whole PVY genomes mentioned above, 166 whole PVY genome sequences were downloaded from the GenBank database (Table 3) to further diversify the dataset for phylogenetic analysis.

2.2. Genome analysis

To further group the 285 genomes into strains or strain variants, sequence alignments were done using either Clustal X or MUSCLE, with some manual adjustment (Larkin et al., 2007; Edgar, 2004). A whole-genome UPGMA tree was generated in RDP4.22 to determine how the isolates clustered together. This information was compared with RT-PCR and serological data when available, and with BLAST data (http://blast.ncbi.nlm.nih.gov/) (Sokal and Michener, 1958; Martin et al., 2010) to classify isolates into strain groups. Using this approach, nearly 96% of the 285 PVY genomes (272/285) were assigned to one of the previously described PVY strain groups (Table 4). There was a slight bias towards PVY^{O} and PVY^{O}-O5 sequences in the combined dataset, due to a larger number of the corresponding whole genome sequences available for analysis. We generally viewed this as an advantage, allowing us a better resolution in the phylogenies of the O-specific sequences present in both recombinant and non-recombinant genomes. Fig. 1 summarizes the recombinant structures revealed in the combined PVY datasets, hereafter the ‘RDP analysis.’ Five “parental” (non-recombinant, potential parent) genomes were assigned to PVY^{O}, $PVY^{Eu\text{-}N}$, PVY^{C}, $PVY^{NA\text{-}N}$, and PVY^{O}-O5 genotypes. Four of these five parental sequences were easily found in recombinant PVY genomes (see Fig. 1), often in multiple recombinant types. However, the majority of the recombinants were found to have PVY^{O} and $PVY^{Eu\text{-}N}$ as parents. One additional undescribed parental sequence type constituted about 75% of the genome for the PVY-NE11 genotype and smaller segments in two additional PVY recombinants (Fig. 1). Only one distinct sequence type, PVY^{O}-O5, was not found in any of the recombinant genomes analyzed (Fig. 1). Seven unusual, or unique, recombinant structures found previously in a very limited number of PVY whole genomes are highlighted on Fig. 1, along with three new structures determined in this study.

Two PVY isolates, NY110001 and AL100001, were identified as novel recombinants between PVY^{C} and PVY^{O} or PVY-NE11 genomes, respectively (Fig. 1). Both are laboratory isolates of unclear origin, which were maintained in tobacco in Ithaca, NY and both displayed an O-serotype (Table 1). When subjected to RT-PCR typing using the protocol of Lorenzen et al. (2006b), both produced a single 267-nt band, indicating a possible PVY^{O} isolate, but when subjected to RT-PCR typing using the protocol of Chikh-Ali et al. (2013), no products were amplified, indicating a new, unknown genotype (not shown). The combination of serological and molecular properties of these two PVY^{C} isolates was very similar to the same properties of a non-recombinant PVY^{C} isolate from tomato described recently (Chikh-Ali et al., 2016c). The two isolates, NY110001 and AL100001, contained a relatively small fragment of the PVY^{O} or PVY-NE11 genome inserted in the CI cistron, respectively. The third isolate found to have a novel structure, ND23, was similar to a typical $PVY^{N:O}$, but with the RJ shifted to the 5′-terminus of the genome (from nt 2390 to nt 2307).

Table 1
Newly sequenced whole-genome isolates used for analyses in this work. For "Origin," a USA state standard two-letter abbreviation is given, with the exception of two isolates from Germany. Strain was determined by a combination of TAS-ELISA ("serology"), two multiplex RT-PCR assays, and phylogenetic and recombination analyses. "Uncl." (unclassified) strains are those which have a novel structure and have not been previously characterized (see Fig. 1). "Tobacco Rxn." (tobacco reaction) describes the symptoms observed 3–4 weeks post-inoculation when the given isolate was mechanically inoculated onto tobacco cv. Burley. M=mosaic; VN=vein necrosis; nd=not determined.

#	Isolate name	Year collected	GenBank accession	Strain	Serology	Multiplex RT-PCR	Tobacco Rxn.	Origin
1	AL100001	Unknown	KY847935	Uncl.	O	Uncl.	M	AL
2	CA14	2005	KY847936	O	O	O	M	CA
3	CO11	2004	KY847937	O5	O5	O	M	CO
4	CO86	2004	KY847938	O	O	O	M	CO
5	ID1_4_32B	2004	KY847939	O5	O5	O	M	ID
6	ID_1_7_12B	2004	KY847940	O	O	O	M	ID
7	ID_1258	2006	KY847941	O	O	O	M	ID
8	ID1_1_3A	2004	KY847942	N-Wi	O	N-Wi	VN	ID
9	ID1_3_11B	2004	KY847943	N-Wi	O	N-Wi	VN	ID
10	ID11_13_11b	2004	KY847944	NTNa	N	NTNa	VN	ID
11	ID11_13_12A	2004	KY847945	N-Wi	O	N-Wi	VN	ID
12	ID12_102IC3	2012	KY847946	NTNa	N	NTNa	VN	ID
13	ID12_110Ban1	2012	KY847947	O5	O5	O	M	ID
14	ID12_22RN8	2012	KY847948	NTNa	N	NTNa	VN	ID
15	ID12_401Chf	2012	KY847949	N-Wi	O	N-Wi	VN	ID
16	ID125	2005	KY847950	N-Wi	O	N-Wi	VN	ID
17	ID1280	2006	KY847951	NE-11	N	NE-11	VN	ID
18	ID13_148Oth	2013	KY847952	N-Wi	O	N-Wi	VN	ID
19	ID13_610Brw	2013	KY847953	NE-11	N	NE-11	VN	ID
20	ID21	2005	KY847954	NE-11	N	NE-11	VN	ID
21	ID26	2005	KY847955	NE-11	N	NE-11	VN	ID
22	ID281_O5	2005	KY847956	O5	O5	O	M	ID
23	ID38	2005	KY847957	NTNa	N	NTNa	VN	ID
24	ID50	2005	KY847958	NTNa	N	NTN	VN	ID
25	ID89	2005	KY847959	N-Wi	O	N-Wi	VN	ID
26	ID90	2006	KY847960	N:O	O	N:O	VN	ID
27	Linda14	2013	KY847961	N-Wi	O	Uncl.	VN	Germany
28	ME_236_120	2004	KY847962	O	O	O	M	ME
29	ME_236_71	2004	KY847963	O	O	O	M	ME
30	ME10	2009	KY847964	NTNa	N	NTNa	nd	ME
31	ME100004	2010	KY847965	N:O	O	N:O	VN	ME
32	ME100007	2010	KY847966	O	O	O	M	ME
33	ME100008	2010	KY847967	O5	O5	O	M	ME
34	ME100011	2010	KY847968	NTNa	N	NTNa	VN	ME
35	ME100031	2010	KY847969	NTNa	N	NTNa	VN	ME
36	ME110008	2011	KY847970	NTNa	N	NTNa	VN	ME
37	ME110032	2011	KY847971	NTNa	N	NTNa	VN	ME
38	ME200cornell	2006	KY847972	O5	O5	O	M	ME
39	ME4	2006	KY847973	NTNa	N	NTNa	VN	ME
40	ME81	2006	KY847974	N:O	O	N:O	M	ME
41	ME9	2005	KY847975	NTNa	N	NTNa	VN	ME
42	MI090004	2009	KY847976	N:O	O	N:O	M	MI
43	MI110011	2011	KY847977	N-Wi	O	N-Wi	VN	MI
44	MN10c_26	2005	KY847978	N:O	O	N:O	M	MN
45	MN121	2006	KY847979	N:O	O	N:O	M	MN
46	MN13a_39	2004	KY847980	N:O	O	N:O	M	MN
47	MN15_G_52	2004	KY847981	N-Wi	O	N-Wi	VN	MN
48	MN21	2005	KY847982	N-Wi	O	N-Wi	VN	MN
49	MN85	2006	KY847983	NTNa	N	NTNa	VN	MN
50	MSU_45-384a	2012	KY847984	Eu-N	N	Eu-N	VN	MT
51	MSU_59-383b	2012	KY847985	Eu-N	N	Eu-N	VN	MT
52	MT100006	2010	KY847986	Eu-N	N	Eu-N	VN	MT
53	MT100010	2010	KY847987	O5	O5	O	M	MT
54	MT100017	2010	KY847988	Eu-N	N	Eu-N	VN	MT
55	MT29	2004	KY847989	O5	O	O	M	MT
56	MT52	2005	KY847990	N:O	O	N:O	VN	MT
57	MT63	2004	KY847991	O5	O	O	M	MT
58	ND100040	2010	KY847992	NE-11	N	NE-11	VN	ND
59	ND110037	2011	KY847993	N:O	O	N:O	VN	ND
60	ND121	2004	KY847994	N:O	O	N:O	M	ND
61	ND18	2005	KY847995	N:O	O	N:O	VN	ND
62	ND2	2004	KY847996	N-Wi	O	N-Wi	VN	ND
63	ND23	2006	KY847997	Uncl.	O	N:O	VN	ND
64	ND35	2004	KY847998	O5	O	O	nd	ND
65	ND65	2005	KY847999	N:O	O	N:O	VN	ND
66	ND68	2004	KY848000	N:O	O	N:O	M	ND
67	ND71	2006	KY848001	N:O	O	N:O	M	ND
68	ND98	2004	KY848002	N-Wi	O	N-Wi	VN	ND
69	ND99	2004	KY848003	N:O	O	N:O	VN	ND
70	NE38	2004	KY848004	O	O	O	M	NE
71	NE40	2004	KY848005	N:O	O	N:O	VN	NE

(continued on next page)

Table 1 *(continued)*

#	Isolate name	Year collected	GenBank accession	Strain	Serology	Multiplex RT-PCR	Tobacco Rxn.	Origin
72	NE6	2005	KY848006	O	O	O	M	NE
73	NY090004	2009	KY848007	N:O	O	N:O	VN	NY
74	NY090029	2009	KY848008	NTNa	N	NTN	VN	NY
75	NY090031	2009	KY848009	O	O	O	M	NY
76	NY100001	2010	KY848010	O	O	O	M	NY
77	NY100002	2010	KY848011	O	O	O	M	NY
78	NY100003	2010	KY848012	O	O	O	M	NY
79	NY100086	2010	KY848013	O	O	O	M	NY
80	NY110001	Unknown	KY848014	Uncl.	O	Uncl.	M	NY
81	NY120001	2012	KY848015	N-Wi	O	N-Wi	nd	NY
82	NY120002	2012	KY848016	N-Wi	O	N-Wi	VN	NY
83	NY51	2006	KY848017	N:O	O	N:O	M	NY
84	OR16	2005	KY848018	N:O	O	N:O	VN	OR
85	OR2	2004	KY848019	O	O	O	M	OR
86	OR20	2004	KY848020	O	O	O	M	OR
87	OR3	2005	KY848021	N-Wi	O	N-Wi	VN	OR
88	OR35	2005	KY848022	NTNa	N	NTNa	VN	OR
89	Pondo4	2013	KY848023	261–4	O	Uncl.	M	Germany
90	SU2	2013	KY848024	NE-11	N	NE-11	VN	MT
91	WA316	2009	KY848025	NE-11	N	NE-11	VN	WA
92	WA70	2005	KY848026	N:O	O	N:O	VN	WA
93	WA9	2005	KY848027	O	O	O	M	WA
94	WI120018	2012	KY848028	NE-11	N	NE-11	VN	WI
95	WI120092	2012	KY848029	O	O	nd	M	WI
96	WI120127	2012	KY848030	N:O	O	N:O	M	WI
97	WI3	2004	KY848031	O	O	O	M	WI
98	WI3406	2006	KY848032	O	O	O	M	WI
99	WI62	2004	KY848033	N:O	O	N:O	VN	WI
100	WY1	2004	KY848034	O5	O	O	M	WY
101	CO_225	2005	KY848035	O5	O5	O	M	CO
102	CO_28	2005	KY848036	O5	O5	O	M	CO
103	CO12	2004	KY848037	O5	O5	O	M	CO
104	CO238	2005	KY848038	O5	O5	O	M	CO
105	CO254	2005	KY848039	O5	O5	O	M	CO
106	CO275	2005	KY848040	O5	O5	O	M	CO
107	CO32	2004	KY848041	O5	O5	O	M	CO
108	CO39	2004	KY848042	O5	O5	O	M	CO
109	CO53	2004	KY848043	O5	O5	O	M	CO
110	CO55	2004	KY848044	O5	O5	O	M	CO
111	CO6	2004	KY848045	O5	O5	O	M	CO
112	ID_1005	2006	KY848046	O5	O5	O	M	ID
113	ID_236	2005	KY848047	O5	O	O	M	ID
114	ME_222_18	2004	KY848048	O5	O5	O	M	ME
115	ME_250_106	2005	KY848049	O5	O5	O	M	ME
116	ME_250_20	2004	KY848050	O5	O5	O	nd	ME
117	ME_323_34	2004	KY848051	O	O	O	M	ME
118	ND127	2004	KY848052	O	O	O	M	ND
119	T1	2007?	KY848053	O	O	O	M	ID

2.3. PVY genome sectioning and phylogenetic analysis

Due to the limited number of RJs in the most common recombinants of PVY, and their relatively conserved positions (see Fig. 1), we decided to use large sections of the PVY genome between these conserved RJ 1–4 positions for phylogenetic analysis. The objective was to see if these genome sections could have originated from different parental sequences within O or N genomic lineages. In this case, we relied on the natural diversity characteristic of PVY^{O} and $PVY^{Eu\text{-}N}$ sequences in non-recombinant as well as in recombinant genomes (Karasev et al., 2011; Ogawa et al., 2012). The entire PVY genome was divided into 5 sections: nt 1–500, 501–2390, 2391–5850, 5851–9200, and 9201–9700, numbered from 1 to 5 (Fig. 1). Phylogenies for sections 2–4 provided better resolution because of their greater length (Figs. 2–4), while trees for sections 1 and 5 were less robust and informative due to being shorter (Supplementary Figs. 1 and 2). Overall, each of the Sections 1–5 allowed clear separation in the phylogenetic trees between the PVY^{O} and PVY^{C} types, and between the $PVY^{Eu\text{-}N}$ and $PVY^{NA\text{-}N}$ types.

2.4. Genome section 2 phylogeny

The sequences in section 2 are predominantly N-type sequences from recombinant genomes (Fig. 1). There was substantial sequence diversity in this region. PVY^{NTNa} and PVY^{E} isolates grouped together, suggesting that the initial N segment of these strains came from a common parent (Fig. 2). Even more tightly related were the $PVY^{N:O}$ isolates, which, other than one isolate (MI090004), formed a single monophyletic group with relatively little intra-strain diversity (99.2% intra-strain identity, compare with Table 5; Fig. 2). $PVY^{NA\text{-}N}$ isolates similarly formed a single distant group with relatively low diversity (Table 5; Fig. 2).

There are other interesting relationships among members of other strains. There were two distinct clades within the PVY-NE11 lineage in Fig. 2. This split can be easily explained because of two slightly different positions (2220 nt vs. 2009 nt) of the RJ separating the 5′-terminal N-sequence and 3′-proximal $PVY^{NA\text{-}N}$ sequence. $PVY^{N\text{-}Wi}$ was easily the most diverse for this section. Despite having a recombinant structure very similar to $PVY^{N:O}$ (Fig. 1), the N segment in $PVY^{N\text{-}Wi}$ did not appear to be related to the N segment in $PVY^{N:O}$. $PVY^{N\text{-}Wi}$ did not form a single clade but rather had one clade plus a number of other

Table 2
Primer set developed and used for whole-genome PVY Sanger sequencing.

Pair #	Forward primer	Forward primer sequence	Reverse primer	Reverse primer sequence	Segment length	Genome span
O-type pairs for full genome sequencing						
1	20startF	CAACATAAGAAAAWCAACGCAAAAAC	18a_1223c	GGCATAYTGTTGRGCACAGGT	1203	20-1223
2	FL5nFn	CAACGCAAAAACACTCAYAAA	18a_1223c	GGCATAYTGTTGRGCACAGGT	1190	33-1223
3	940F	GGAGACTTGTTCATAGTGCGTGG	2076R	GTTAATGTAACAGAAGCCTTGCCTGG	1136	940-2076
4	1512F	GCTCGARYTAGCAAGGTTCCAGAAG	2323R	GATATCCAGTYGTYTGYGAGCC	811	1512-2323
5	1868F	RRAARTGCACGAGTTCRAAAGATGG	2500R	GTTTCAGYGCTGCCGACTCAGAC	632	1868-2500
6	2299F	GAAAGCATCTAGCGTGTCCCAAC	3074R	CTTTCAAATCTGCCTTTCCTGTGGG	775	2299-3074
7	2769F	CCTTGATGCTACGTGYGATGGRTTC	3517R	CTGGTGTGGARCGCTGGTGTC	748	2769-3517
8	3283F	CGTAGTGGCAGTGTGTCAGGC	4304R	GTTGTRAATTCRACCTCTCTTCCCAC	1021	3283-4304
9	3950F	GTTGCCTGTTCATCTTAGTGTAGCC	4969R	GGGCAACGAAATTCTGGATAAAG	1019	3950-4969
10	4839F	CTTRCCAGTGATGACAGGAGGCG	5568R	CGCAGCTTGGTGATGAACGAAC	729	4839-5568
11	5760F	CTCGTGACAAAAGGGCTGG	28_o6505c	CTCTCATTCTGCACTGGAGC	745	5760-6505
12	6199F	GAGAGAGARCTCGAAYTAAGGCARAC	7194R	CCCATACGCATCCATCAAAGGC	995	6199-7194
13	14_o7008	GAGCAAGCTAAGCACTCTGC	8111R	GTCAAGAATGCCCTCTTTATCCG	1103	7008-8111
14	7739F	GGGTATACTGTGATGCTGATGGC	CPBC	ACGCTTCTGCAACATCTGAG	1336	7739-9075
15	YFL2_8567f	GCAAATGACACAATTGATG	FL_12R_9700R	TTTTTTTGTCTCCTGATTGAAGTTTACAG	1133	8567-9700
16	PVY100_4FP	TTGACTTTTATGAGGTCACATCACG	FL3new	TTTTTTGTCTCCTGATTGAAGTTTACAGTCAC	535	9165-9700
N-type pairs for full genome sequencing						
17	20startF	CAACATAAGAAAAWCAACGCAAAAAC	1216R	GTTGRGCACAGGTRGGGCAGG	1196	20-1216
18	20startF	CAACATAAGAAAAWCAACGCAAAAAC	1142R	GCCACAGTCTTCAACTGGTAAGCC	1122	20-1142
19	FL5nFn	CAACGCAAAAACACTCAYAAA	1216R	GTTGRGCACAGGTRGGGCAGG	1183	33-1216
20	FL_2F_800F	CAAATGGTCTAATCAAGTCCGCAC	2076R	GTTAATGTAACAGAAGCCTTGCCTGG	1276	800-2076
21	841F	CAAATGGTMTAATCAAGTCCGCAC	1887R	GTGCATTTTCTGCTGACTCCTGG	1046	841-1887
22	1864F	GAAAATGCACGAGTTCGAAAGATGG	2554R	CATCYARTAGYAAYTGYYTCATCAC	690	1864-2554
23	2274F	CAGACATGCCATGTGGTTGACTCG	3515R	GGTGTGGAGCGCTGATGYCG	1241	2274-3515
24	3118F	GCACCGCCTCAGGGTTRAATG	4390R	GCATCAACAAATGATTGGAAAGAC	1272	3118-4390
25	3909F	GAGGGGAGCTGTTGGGTCTGG	4969R	GGGCAACGAAATTCTGGATAAAG	1060	3909-4969
26	4756F	GCGTATTGGRCACACCGAAAAGG	5578R	CGAGTGMAGTTGTTGCTTGATGATG	822	4756-5578
27	5024F	CGCCCTTGTGTGATCAATCCATACC	5919R	CATGTTGATGAACTTCCTGCTTGAC	895	5024-5919
28	5554F	CATCATCAAGCAACAACTKCACTCG	6564R	GAAAACAGGGAARTCCTTTGGCATC	1010	5554-6564
29	6271F	GTGGAGCATGAAGCCAAATCACTC	7523R	GATCCATTCCATATRCCAAGYGAG	1252	6271-7523
30	7142F	CTGTGGATTCGGAAGCAGAAGC	8111R	GTCAAGAATGCCCTCTTTATCCG	969	7142-8111
31	15_p7498	GCTTGGCATATGGAATGGAT	8720R	GCTTTAATTCGTGGCACAGTRTGAG	1222	7498-8720
32	YFL2_8567f	GCAAATGACACAATTGATG	CPBC	ACGCTTCTGCAACATCTGAG	508	8567-9075
Additional pairs for full genome sequencing						
33	842F	CAAATGGTCTAATCAAGTCCGCAC	2163R	CTTTGGCACACACATGTCACGAAC	1321	842-2163
34	872F	CATACGAAGGGGTGATAGTGGAGTC	2179R	GGTTCCAAGCTTCGGCACACAC	1307	872-2179
35	987F	CTATGATGCRCGTTCCAAGGTTACDC	2076R	GTTAATGTAACAGAAGCCTTGCCTGG	1089	987-2076
36	1015F	GGRGTTMTRGAYTCAATGGTTCAG	2076R	GTTAATGTAACAGAAGCCTTGCCTGG	1061	1015-2076
37	2299F	GAAAGCATCTAGCGTGTCCCAAC	3570R	CGAAAACCATGATGACTAAAGCC	1271	2299-3570
38	3118F	GCACCGCCTCAGGGTTRAATG	4388R	CTACGTCGGCATTGGTTTTTGAGC	1270	3118-4388
39	3283F	CGTAGTGGCAGTGTGTCAGGC	4390R	GCATCAACAAATGATTGGAAAGAC	1107	3283-4390
40	3950F	GTTGCCTGTTCATCTTAGTGTAGCC	PVY100_11RP	GTTATATGCAAAGCAAGCAAGAGC	885	3950-4835
41	4839F	CTTRCCAGTGATGACAGGAGGCG	5919R	CATGTTGATGAACTTCCTGCTTGAC	1080	4839-5919
42	6134F	GATTTGATGCCACACAACCCACTC	7061R	CTTGCAAATTCCCTGTTAGAGCC	927	6134-7061
43	5760F	CTCGTGACAAAAGGGCTGG	7194R	CCCATACGCATCCATCAAAGGC	1434	5760-7194
44	5760F	CTCGTGACAAAAGGGCTGG	7304R	GCCTCTTCAAAAGCATCACAGTC	1544	5760-7304
45	6688F	GCAGCACATTCTGGAARCATTGG	7483R	GACAGCTTTGCATGACTATTTCTTCC	795	6688-7483
46	6688F	GCAGCACATTCTGGAARCATTGG	8239R	GYAYGTACATGCCCTCAATCAG	1551	6688-8239
47	7700F	CTAAGTTTTATGGAGGTTGGGACAG	8957R	CCGTTGATGTTTGGCGAGGTTC	1257	7700-8957
48	7739F	GGGTATACTGTGATGCTGATGGC	8423R	CCCTTCCTGCGCTATTGTTGC	684	7739-`8423

isolates in other locations on the tree (Fig. 2). This higher diversity of $PVY^{N\text{-}Wi}$ is also reflected in the intra-strain identity percentage, which at 98.2% is the lowest of all the recombinants (Table 5).

Several other recombinants were similar to $PVY^{N:O}$ and $PVY^{N\text{-}Wi}$, in that they had similar genome structure, but different parental N sequences. PVY^{NTNa} and PVY^{NTNb} isolates were phylogenetically unrelated to one another for section 2, with PVY^{NTNb} forming a clade by itself. Even within a specific recombination pattern, such as PVY-SYR-II, there was an apparent diversity of clade placement – and thus parental sequence – for different isolates. Some PVY-SYR-II isolates shared a parent of their section 2 with PVY-SYR-I (such as 1103 and 1108), some with unclassified isolates (such as Wilga156 and Wilga156var), and some with none of the isolates included in these analyses, such as SYR_II_Be1 (Fig. 2). All of these observations suggested that these recombinant structures could have arisen more than once from different parents providing specific fragments for their section 2. This conclusion appeared to be supported by the parental strain $PVY^{Eu\text{-}N}$, which also did not form one tight clade. It seemed that diversity within parental strains PVY^{O} and $PVY^{Eu\text{-}N}$ led to even further diversity within recombinant strains, despite similar structures arising repeatedly.

The section 2 area represented almost exclusively non-recombinant sequences in the O-type genomes, and thus gave the best picture of PVY^{O} sequence diversity between genome positions 501 and 2390 for all non-recombinant PVY^{O} genomes (Fig. 2). Eight O-specific clades could be seen within the PVY^{O} lineage, with five additional major clades that included PVY^{O}-O5 sequences (Fig. 2). Isolate T1 appeared to be unique; it grouped closer to the PVY^{O}-O5 isolates but stood alone among isolates of the PVY^{O}-O5 clade, although it has historically been considered a PVY^{O} strain isolate (Fig. 2). Repeated and extensive analyses confirmed it as a non-recombinant.

The PVY^{C} strain is quite diverse (only 90.5% strain identity; Table 5), as seen immediately by the long branch lengths and lack of any ability to group them into clades (Fig. 2). As a group, they are more closely related to the O-type isolates than the N-type isolates for this section (Fig. 2).

Table 3
Whole genome sequences from GenBank used for analyses in this work. Strain was determined by phylogenetic and recombination analyses; GenBank strain designations were disregarded. "Unclassified" strains are those which have a novel structure (see Fig. 1).

#	Isolate name	GenBank accession	Strain
1	1101	KC296434	SYR-I
2	1103	KC296435	SYR-I
3	1104	KC296436	Unclassified
4	1105	KC296437	NTNa
5	1106	KC296438	SYR-II
6	1107	KC296439	Unclassified
7	1108	KC296440	SYR-II
8	11439	KC634005	NTNa
9	09_3a	JF795485	N-Wi
10	11227_2	KC634004	NTNa
11	11627_12	KC634007	NTNa
12	12_94	AJ889866	NTNb
13	261_4	AM113988	261-4
14	34_01	AJ890342	NTNb
15	423_3	AY884982	NTNa
16	9703_3	KC296432	Unclassified
17	9703_4	KC296441	NTNa
18	9703_5	KC296433	SYR-I
19	A95	HQ912866	N:O
20	Adgen	AJ890348	C
21	AGA	JF928459	E
22	ALF_VI	JQ924287	NTNa
23	Alt	AY884985	N:O
24	AQ4	JN083841	N-Wi
25	AST	JF928460	NTNa
26	CO1750	HQ912910	O5
27	CO1801	HQ912898	O5
28	CO1827	HQ912912	O5
29	CO1898	HQ912906	O5
30	CO1960	HQ912915	O5
31	CO2081	HQ912913	O
32	CO2122	HQ912897	O
33	CO2140	HQ912914	O
34	CO2146	HQ912907	O5
35	CO2194	HQ912901	O5
36	CO2247	HQ912899	O5
37	CO2272	HQ912900	O5
38	CO2294	HQ912903	O5
39	CO2352	HQ912902	O5
40	CO2374	HQ912908	O5
41	CO284	HQ912905	O5
42	CO286	HQ912911	O5
43	CO289	HQ912904	O5
44	CO303	HQ912909	O5
45	CW	HQ912865	O
46	Del_66	JN034046	N-Wi
47	Ditta	AJ890344	NTNa
48	E30	HM991453	Unclassified
49	Eu_12Jp	AB702945	NTNa
50	FL	HM367075	O
51	FrKV15	HM991454	Unclassified
52	FZ10	JN083842	Unclassified
53	GBVC_PVY_10	JQ969036	Eu-N
54	GBVC_PVY_15	JQ969034	NTNa
55	GBVC_PVY_23	JQ969040	Unclassified
56	GBVC_PVY_26	JQ969039	N-Wi
57	GBVC_PVY_3	JQ969035	NTNa
58	GBVC_PVY_34	JQ969041	N-Wi
59	GBVC_PVY_9	JQ969037	NTNb
60	Gr99	AJ890343	NTNb
61	HC_2quan	HM590406	NTNb
62	HN1	HQ631374	NTNa
63	HN2	GQ200836	SYR-I
64	HR1	FJ204166	NTNa
65	Hun_NTN	M95491	NTNa
66	ICIA	HQ912864	O
67	ID1_5_62A	HQ912890	O
68	ID1010	HQ912887	O5
69	ID11_27_57B	HQ912885	O5
70	ID1269	HQ912882	O5
71	ID130	HQ912888	O
72	ID14_2_14A	HQ912870	N:O
73	ID155	HQ912869	NTNa
74	ID20	HQ912867	NE-11
75	ID243	HQ912895	O
76	ID253	HQ912880	O5
77	ID269	FJ643477	O5
78	ID281	HQ912893	O
79	ID315	HQ912881	O5
80	ID331	HQ912879	O5
81	ID431	HQ912862	N:O
82	ID883	HQ912894	O
83	ID968	HQ912886	O5
84	ID988	HQ912883	O5
85	IUNG_11	JF927759	NTNa
86	IUNG_13	JF927761	NTNa
87	IUNG_15	JF927763	NTNa
88	IUNG_3	JF927751	N-Wi
89	IUNG_4	JF927752	NTNa
90	IUNG_5	JF927753	N-Wi
91	IUNG_7	JF927755	261-4
92	IUNG_9	JF927757	NTNa
93	L26	FJ204165	NTNa
94	L56	AY745492	N:O
95	Linda	AJ890345	NTNa
96	LR	HQ912896	N-Wi
97	LW	AJ890349	N-Wi
98	M3	KF850513	NTNa
99	MAF_VOY	JQ924286	N-Wi
100	Mb112	AY745491	N:O
101	ME120	HQ912892	O
102	ME131	HQ912874	O5
103	ME142	HQ912871	N:O
104	ME162	HQ912872	N:O
105	ME162_CN	JQ971975	NE-11
106	ME173	FJ643479	O
107	ME178	HQ912875	O5
108	ME200	HQ912889	O
109	ME227	HQ912877	O5
110	ME236_4	HQ912891	O
111	ME236_77	HQ912873	O5
112	ME27	HQ912878	O5
113	ME286_58	HQ912884	O5
114	ME56	FJ643478	O5
115	ME89_107	HQ912876	O5
116	MON	JF928458	E
117	Mont	AY884983	Eu-N
118	MV175	HE608964	N-Wi
119	MV99	HE608963	N-Wi
120	N_Egypt	AF522296	Unclassified
121	N_JG	AY166867	NA-N
122	N_Nysa	FJ666337	Unclassified
123	N1	HQ912863	N-Wi
124	N3	HQ912868	N-Wi
125	N4	FJ204164	NTNa
126	N605	X97895	Eu-N
127	NC57	DQ309028	C
128	NE_11	DQ157180	NE-11
129	Nicola	AJ890346	Unclassified
130	NN300_41	JN936422	NA-N
131	nnp	AF237963	Unclassified
132	NTND6	AB331515	NA-N
133	NTNHO90	AB331517	NA-N
134	NTNKGAM1	AB711144	NA-N
135	NTNNN99	AB331518	NA-N
136	NTNOK105	AB331516	NA-N
137	NTNON92	AB331519	NA-N
138	NZ	AM268435	Eu-N
139	O_139	U09509	O5
140	Oz	EF026074	O
141	PB209	EF026076	N:O
142	PB312	EF026075	NTNa
143	PRI_509	EU563512	C
144	PVY_Fr	D00441	Unclassified
145	PVYOUK	JX424837	O
146	RB	HM367076	O5

(continued on next page)

Table 3 *(continued)*

#	Isolate name	GenBank accession	Strain
147	RRA_1	AY884984	NA-N
148	SASA_61	AJ585198	NA-N
149	SCRI_N	AJ585197	Unclassified
150	SCRI_O	AJ585196	O
151	SD1	EU182576	Unclassified
152	SGS_AG	JQ924288	N-Wi
153	SON41	AJ439544	C
154	SYR_II_2_8	AB461451	SYR-II
155	SYR_II_Be1	AB461452	SYR-II
156	SYR_II_DrH	AB461453	SYR-II
157	SYR_III_L4	AB461454	SYR-III
158	SYR_NB_16	AB270705	SYR-I
159	T13	AB714135	Unclassified
160	Thole	M95491	NTNa
161	Tu660	AY166866	NA-N
162	v942490	EF016294	NTNa
163	Wilga156	AJ889867	SYR-II
164	Wilga156var	AJ889868	Unclassified
165	Wilga5	AJ890350	N-Wi
166	YO_ANT	JQ924285	O

Table 4
Breakdown of all strains used for this study, including both newly sequenced isolates and GenBank isolates.

Strain	# Newly sequenced	# from GenBank	Total #
O	23	19	42
O5	27	35	62
N-Wi	16	16	32
N:O	22	9	31
NTNa	15	28	43
NTNb	0	5	5
Eu-N	4	4	8
NA-N	0	11	11
E	0	2	2
C	0	4	4
SYR-I	0	5	5
SYR-II	0	6	6
SYR-III	0	1	1
NE-11	8	3	11
261-4-like	1	1	2
Unclassified	3	17	20
Total #	119	166	285

2.5. Genome section 3 phylogeny

Unlike section 2, most of the sequences analyzed for section 3 represented O-type sequences from both recombinant and non-recombinant PVY genomes (see Fig. 1). Most recombinant PVY isolates carrying O-sequences in this section, including PVY^{NTNa}, PVY^{NTNb}, PVY^{E}, $PVY^{N:O}$, $PVY^{N\text{-}Wi}$, and PVY-SYR-I to -III types, grouped separately from the non-recombinant sequences analyzed (Fig. 3), although some $PVY^{N\text{-}Wi}$, PVY^{NTNb}, and PVY-SYR-I and -II isolates were related to non-recombinant O-type clades, for example PVY^{NTNb} isolate Gr99, $PVY^{N\text{-}Wi}$ isolate N3, and PVY^{O} isolate YO_ANT (Fig. 3). This again suggested that these structures arose multiple times, i.e., these strains are polyphyletic in origin.

All PVY^{NTNa} isolates were found in a single lineage comprising only recombinant types PVY^{NTNa} and PVY^{E} (Fig. 3). Interestingly, none of the non-recombinant PVY^{O} isolates analyzed were related to the PVY^{O} parent providing the PVY^{NTNa} section 3 (see Fig. 3), hence no non-recombinant PVY^{O} sequences used for this study were present in the recombinant lineage with PVY^{NTNa} isolates. However, some of the Syrian and PVY^{NTNb} isolates, such as SYR_II_2_8 and 34_01, respectively, had a section 3 parent closer to the parent of the PVY^{NTNa} section. Both PVY^{E} isolates (MON and AGA) were found in the same lineage with PVY^{NTNa} isolates and hence likely shared a parent with PVY^{NTNa} for section 3.

All 31 $PVY^{N:O}$ isolates were monophyletic for this section, forming a distinct, single clade (Fig. 3), and no non-recombinant PVY^{O} isolates were closely related to the hypothetical PVY^{O} parent of the $PVY^{N:O}$ section 3. This was consistent with the result for $PVY^{N:O}$ for section 2 except that the one isolate which previously grouped separately (MI090004) now grouped with the rest, making a truly monophyletic group for $PVY^{N:O}$ section 3.

Interestingly, once again no PVY^{O}-O5 segments were found in any of the recombinant PVY isolates; instead they all grouped into the same 5 clades as for section 2 (Fig. 3). This indicated that although the PVY^{O} and PVY^{O}-O5 strains are closely related, PVY^{O}-O5 was never seen as a relative to the parent for any recombinant O-type sequences for section 3. PVY^{C}, on the other hand, is a much more distant relative of the other strains. PVY^{C} isolates again had long branch lengths and could not be grouped into meaningful clades, but like PVY^{O}-O5 were, as a whole, closer to O-type isolates than N-type isolates.

For section 3, all N-types were clearly clustered into separate clades: $PVY^{NA\text{-}N}$, PVY-NE11, and $PVY^{Eu\text{-}N}$ (Fig. 4). PVY-NE11 sequences were split into the same two major clades (Fig. 3) as for section 2 (Fig. 2). In this case, however, it could not be related to differences in the positions of the corresponding recombinant junctions, but rather reflected actual evolutionary relationships between these isolates (Fig. 3). $PVY^{NA\text{-}N}$ and $PVY^{Eu\text{-}N}$ isolates formed nearly the exact same clades as for section 2 (Fig. 2).

2.6. Genome section 4 phylogeny

Section 2 contained mostly N-types and section 3 contained mostly O-types, but section 4 was roughly an even split between O- and N-types (Fig. 1). The O-type sequences of section 4 (see Fig. 1) were from non-recombinant PVY^{O} isolates and recombinant types $PVY^{N:O}$ and $PVY^{N\text{-}Wi}$. The $PVY^{N:O}$ recombinant type was again found in its own distinct clade, but this separate clade was clustered with a larger lineage combining recombinant and non-recombinant PVY^{O}. Some sequences from one PVY^{O} clade (e.g. isolate WI3) were placed in a lineage including both $PVY^{N:O}$ and $PVY^{N\text{-}Wi}$ recombinants (Fig. 4). The other non-recombinant PVY^{O} clades did not include recombinant PVY isolates (Fig. 4). The PVY^{O}-O5 sequences were still grouped in a single large lineage of five PVY^{O}-O5 clades, and again no recombinants were found close to PVY^{O}-O5 sequences. The PVY^{C} lineage for section 4 remained similar to the PVY^{C} lineage for the previous two sections: long branch lengths, high diversity, no discernable clades, and more similar to O-type sequences than N-type sequences.

Note that isolate Wilga156var was intentionally left out of this phylogeny because it contains a recombinant breakpoint between an O-type segment and an N-type segment nearly in the middle of this section, and thus its placement would have only reflected the structure, not the actual origins of its section 4 (Fig. 1). Isolate SD1 has an as yet unclassified structure, but due to its placement somewhat between the O-types and N-types it likely has a similar mixed structure for this section (Fig. 4).

The N-type sequences of section 4 were from non-recombinant $PVY^{Eu\text{-}N}$, $PVY^{NA\text{-}N}$, and various recombinants including PVY-NE11, PVY^{NTNa}, PVY^{NTNb}, and PVY-SYR-I to -III (see Fig. 1). The PVY^{NTNa} isolates again formed a single lineage, which included all PVY^{NTNb} isolates (Fig. 4). Interestingly, all Syrian isolates were placed in their own separate lineage, suggesting that although their N-type parents for section 2 (Fig. 2) had non-Syrian relatives, their N-type segment for section 4 (Fig. 4) was likely unique to Syrian recombinants (see Fig. 1). The non-recombinant N-type isolates formed two distinct clades, $PVY^{NA\text{-}N}$ and $PVY^{Eu\text{-}N}$, with no recombinant isolates nearby or mixed in with them (Fig. 4). The PVY-NE11 isolates grouped into the same two lineages as for sections 2 and 3, although now both PVY^{E} isolates grouped near one of the PVY-NE11 clades, namely the one with the shorter $PVY^{NA\text{-}N}$ segment (Fig. 4). The non-recombinant $PVY^{NA\text{-}N}$

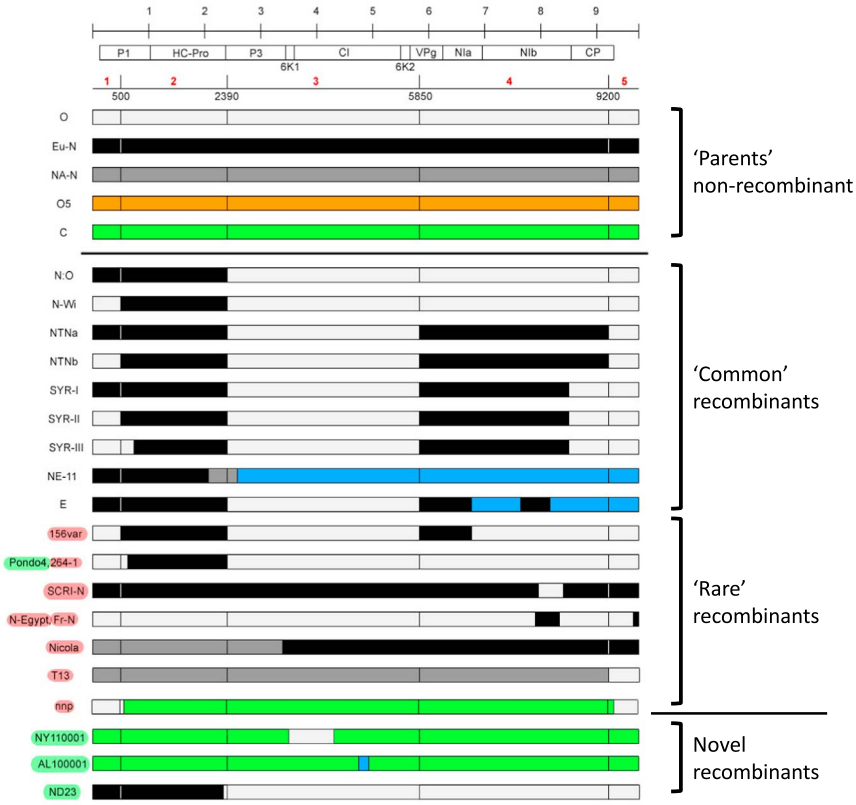

Fig. 1. A schematic diagram of previously known and published *Potato virus Y* (PVY) recombinant structures, along with the three new ones described in this study. The ruler at the top represents the PVY genome (ca. 9.7-kb); individual cistrons are presented as rectangles below that with corresponding protein names. Potential parental sequences are grouped above the upper black horizontal dividing bar, with different parents colored differently. Recombinants are below the upper black bar, and fragments originating from different parents are colored accordingly. The blue parent (e.g. the 3′ section of PVY-NE11) is an as yet unknown strain which has never been found as a non-recombinant. The five major recombinant sections, based on common breakpoints, are designated in red numbers and the locations of breakpoints are given below that. The novel structures determined in this study (NY110001, AL100001, and ND23) are below the lower horizontal black bar. Isolate names highlighted in green indicate isolates sequenced as part of this study which have rare or unique recombinant structures. Those highlighted in red were retrieved from GenBank and found to have rare or unique recombinant structures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

isolates formed just a single lineage, separate from other N-types other than the unclassified recombinants such as isolate T13. Overall, the section 4 analysis continued the pattern of the previous sections, where similarity of recombinant structure did not necessarily correlate with similarity of origin of each N-type or O-type genome segment.

3. Discussion

Recombination, reassortment, and accumulation of mutations are the main forces shaping the evolution of positive strand RNA viruses (Simon and Bujarski, 1994; Roossinck, 2003; Nagy and Simon, 1997; Nagy, 2008), with recombination being one of the main factors in the evolution of potyviruses (Gibbs and Ohshima, 2010). For PVY, the occurrence of multiple recombinant structures is well established (Glais et al., 2002; Lorenzen et al., 2006a, 2006b; Ogawa et al., 2008, 2012; Hu et al., 2009a, 2009b; Karasev and Gray, 2013a, 2013b). Currently, recombinant strains of PVY are prominent or dominant among PVY isolates circulating in potato in Europe, North America, South America, Asia, and Africa (Blanchard et al., 2008; Chikh-Ali et al., 2016a, 2016b; Djilani-Khouadja et al., 2010; Funke et al., 2017; Gray et al., 2010; Galvino-Costa et al., 2012b; Schubert et al., 2015; Visser et al., 2012). The reasons for the emergence and dispersal of multiple recombinants of PVY are not completely understood, and complex interactions between the virus and various hosts supporting replication of PVY may be at least partially responsible for the spread of PVY recombinants (Karasev and Gray, 2013a; Funke et al., 2017). Difficulties in reconstruction of PVY strain evolution and tracing the origin of recombinants can be explained by the extensive international trade in potato seed that may move PVY over long distances and between multiple locations.

Since the number of PVY recombinants circulating in nature was found to be relatively limited (Hu et al., 2009b), the question was posed: how often do the recombination events between different PVY strains actually occur? Specifically, if these were relatively rare events, it would have been possible to reconstruct the pathway of emergence of all main types of PVY recombinants (Karasev et al., 2011; Ogawa et al., 2012), and even provide approximate dates of their emergence (Visser et al., 2012; Gibbs et al., 2017). However, phylogenetic reconstructions of the origins of PVY recombinants are complicated by the necessity to account for the recombination events (Lorenzen et al., 2008; Karasev et al., 2011; Ogawa et al., 2012; Visser et al., 2012; Quenouille et al., 2013). One of the ways to avoid this is to exclude all recombinants from the analysis (Moury, 2010; Cuevas et al., 2012; Chikh-Ali et al., 2016c), but in this case the whole point of the analysis will be lost. There are additional ways to deal with recombination. One would be to exclude the tracts of sequence inherited from the parent that contributed the smaller fraction of the recombinants' genomes and another would be to decompose the recombinants into their constituent parts (making each part a separate sequence in the alignment) and then construct trees containing all the bits and pieces of recombinant sequences as separate sequences. Another way to deal with this recombination problem is to use only partial sequences of the PVY genome, between main RJs (Karasev et al., 2011; Ogawa et al., 2012). Indeed, this approach revealed that different types of PVY recombinants might have arisen different numbers of times, with $PVY^{N:O}$ recombinants found to comprise respective isolates in a single, perhaps monophyletic clade,

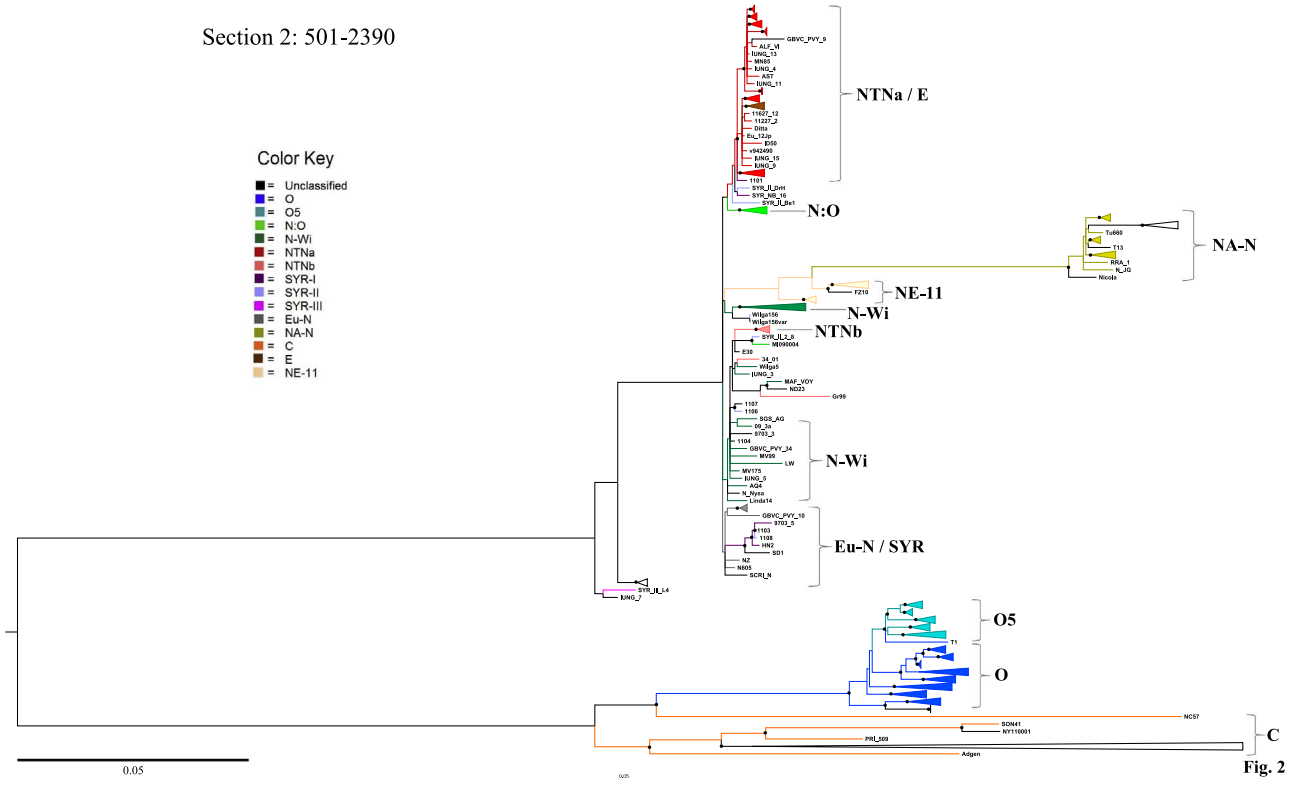

Fig. 2. Phylogenetic tree for section 2, nucleotides 501–2390 (see Fig. 1). Different strain types are colored differently according to the key, which was determined during the course of this study and is based on both phylogenetic and recombination analyses, as well as Fig. 1. The scale bar is the number of substitutions per position (e.g. a scale bar of 0.05 means 5 substitutions per 100 nucleotide positions). Black dots indicate nodes with at least 70% bootstrap support. Collapsed branches contain only isolates of the color-indicated strain type. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

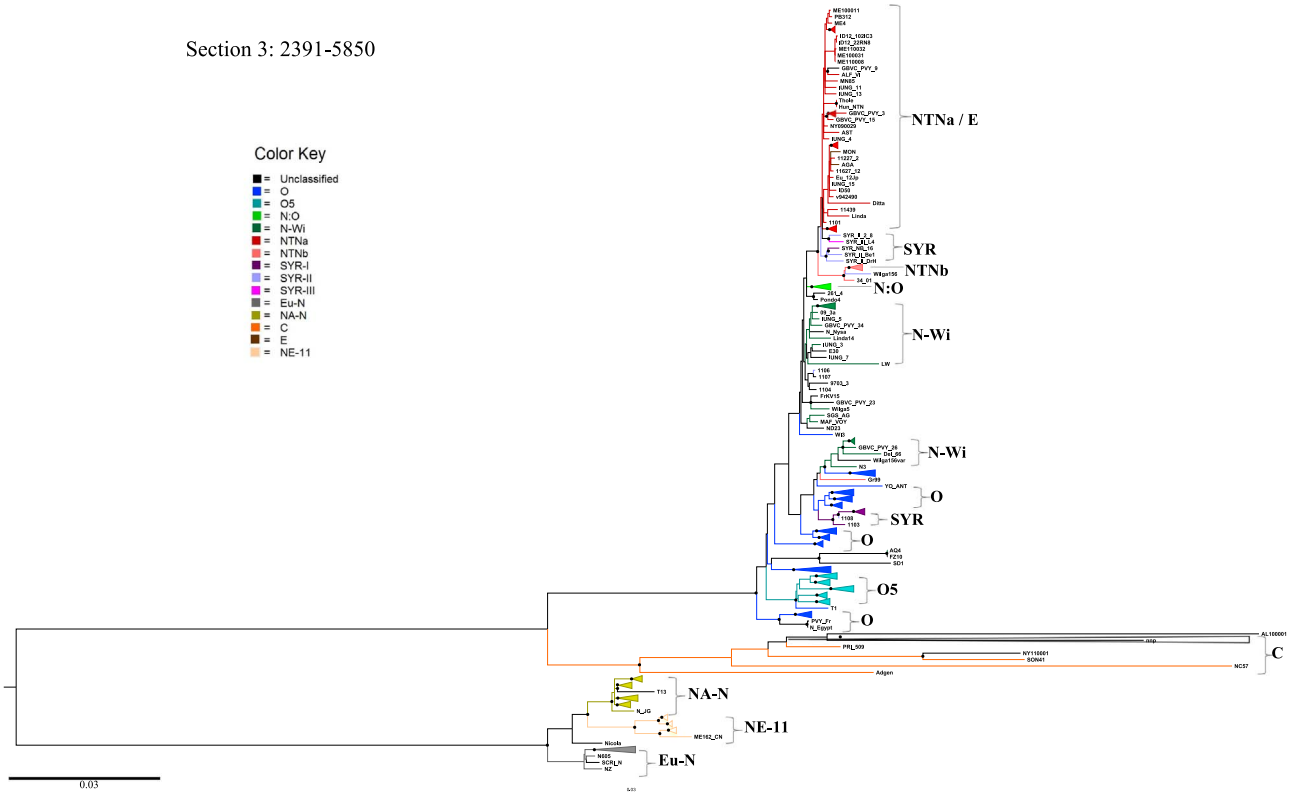

Fig. 3. Phylogenetic tree for section 3, nucleotides 2391–5850 (see Fig. 1). Different strain types are colored differently according to the key, which was determined during the course of this study and is based on both phylogenetic and recombination analyses, as well as Fig. 1. The scale bar is the number of substitutions per position (e.g. a scale bar of 0.03 means 3 substitutions per 100 nucleotide positions). Black dots indicate nodes with at least 70% bootstrap support. Collapsed branches contain only isolates of the color-indicated strain type. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

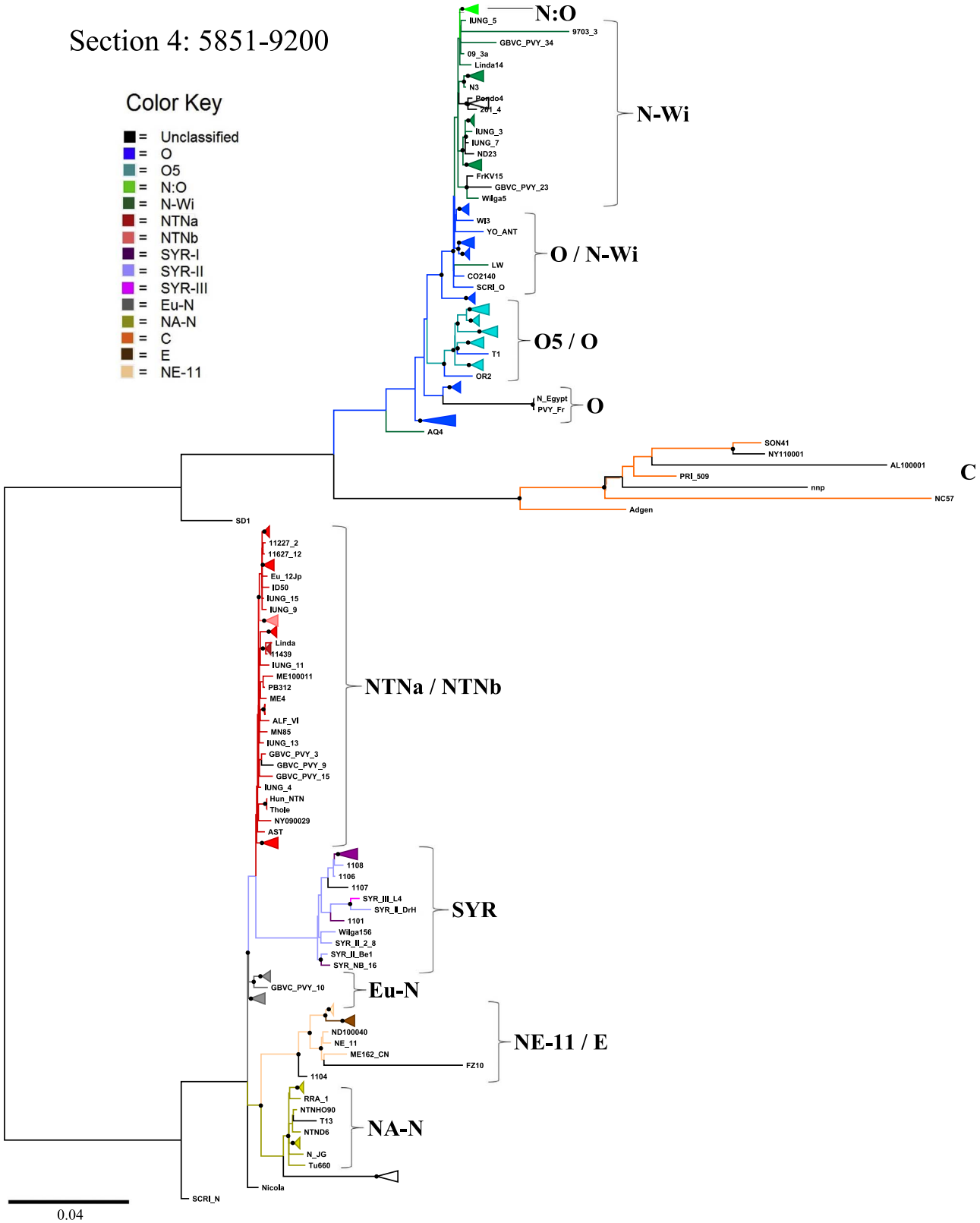

Fig. 4. Phylogenetic tree for section 4, nucleotides 5851–9200 (see Fig. 1). Different strain types are colored differently according to the key, which was determined during the course of this study and is based on both phylogenetic and recombination analyses, as well as Fig. 1. The scale bar is the number of substitutions per position (e.g. a scale bar of 0.04 means 4 substitutions per 100 nucleotide positions). Black dots indicate nodes with at least 70% bootstrap support. Collapsed branches contain only isolates of the color-indicated strain type. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

while PVY^{N-Wi} recombinants were found in multiple clades (Karasev et al., 2011). However this analysis included only one section of the PVY genome, between RJs 2 and 3 (Karasev et al., 2011), or was conducted on a small number of whole PVY genomes (Ogawa et al., 2012; Visser et al., 2012).

Here, the phylogenetic analysis was extended to all five individual

Table 5.
Summary of average percent of intra- and inter-strain identity for whole genomes used in this study. Inter-strain identities for recombinants have been left out because the recombinant structures inherently make the results uninformative, as they will simply be weighted averages of the average percent identities of the recombinant parents of each recombinant section.

# Isolates	Strain	O	Eu-N	NA-N	O5	C	N:O	N-Wi	NTNa	NE-11
42	O	98	82.6	83.1	97.2	88.6	–	–	–	–
8	Eu-N	x	98.8	95.7	82.3	82.2	–	–	–	–
11	NA-N	x	x	99	82.3	82.5	–	–	–	–
62	O5	x	x	x	99	88.4	–	–	–	–
4	C	x	x	x	x	90.5	–	–	–	–
31	N:O	x	x	x	x	x	99.2	–	–	–
32	N-Wi	x	x	x	x	x	x	98.2	–	–
43	NTNa	x	x	x	x	x	x	x	99.1	–
11	NE-11	x	x	x	x	x	x	x	x	98.6

sections of the PVY genome between the four main recombinant junctions (see Fig. 1), and substantially expanded the number of analyzed PVY genomes, which included 119 newly sequenced genomes and 166 whole genomes from GenBank. These sequences represented all types of PVY recombinants and can be considered a relatively unbiased set of PVY sequences suitable for a global analysis of PVY recombinants. Three questions that were addressed were the same posed before (Karasev et al., 2011), with some slight modifications: i) did all similar recombinant types originate from the same parental sequences? ii) do some recombinants represent intermediates between other recombinant types? iii) how often do the recombination events happen between PVY strains?

Examination of the phylogenetic trees presented in Figs. 2–4 suggested that there might be no monophyletic lineages of PVY recombinant structures. Phylogenies of sections 1 and 5 (see Supplementary Figs. 1 and 2) were consistent with this conclusion but provided much less resolution due to their shorter length. Even the $PVY^{N:O}$ lineage comprising 30 or 31 corresponding sequences out of 32 was not monophyletic, with 2 (Fig. 2) or 1 (Fig. 3) additional clades comprised of MI090004 and ND23 isolates. The ND23 isolate was found to be unusual, however, having the RJ2 shifted in the 5′ direction relative to a typical $PVY^{N:O}$, to nt 2307 from nt 2390 position (see Fig. 1). Thus it was considered “unclassified” despite being very similar to a typical $PVY^{N:O}$ in structure.

The answer to the second question may be easier to obtain this time, since at least one strain of PVY, PVY^{E}, was found to represent a recombinant with two other recombinants identified as parents, PVY^{NTN} and PVY-NE11 (Galvino-Costa et al., 2012a). PVY^{E}, thus, is a strain which is a recombinant of recombinants, and so at least once this type of event has occurred in nature.

The answer to the third question, about the frequency of the recombination events between PVY strains, is more complicated. On one hand, among the 119 whole genomes sequenced in this work, only 3 novel recombinant types were found: ND23 mentioned above that may be termed an atypical $PVY^{N:O}$, and two PVY^{C} recombinants, NY110001 and AL100001, with novel recombinant structures (Fig. 1). Isolate AL100001 represents a very unusual recombinant between PVY^{C} and PVY-NE11 sequences that was never reported before. On the other hand, the presence of certain recombinants in multiple clades, either with (PVY^{N-Wi}) or without (PVY^{NTNa}) non-recombinant isolates, suggested that the same types of recombinants were formed more than once from different parental sequences. Hence, based both on the large number of recombinant types of PVY, nineteen found in this work (Fig. 1), and also on the multiple clades characteristic of the same or similar recombinant types (Figs. 2–4), we conclude that recombination between different strains of PVY is relatively frequent. Nevertheless, another conclusion would be that the types of recombinants that result from these recombination events may be relatively limited or restricted. In other words, there is a possibility of selection favoring the survival of particular recombinant structures over others: i.e. that the "common" structures which appear to have independently arisen multiple times could simply be the small highly fit subset of a much larger pool of recombinant genomes that arise in nature but which are predominantly less fit than the parental genomes from which they are derived. The nature of such a limitation or restriction was demonstrated to be unrelated to the physical properties of the PVY RNA genome around the most common RJs (Hu et al., 2009b), and would be more likely related to some form of selection pressure provided by the host, perhaps expressing various forms of resistance to the virus.

4. Materials and methods

4.1. Virus sources, RNA extraction and RT-PCR amplification

Of the 119 PVY isolates that were sequenced for this study, 107 came from a national PVY survey conducted in the United States between 2004–2006; this survey and collection methodology were described elsewhere (Gray et al., 2010). Specific locations of the isolate collection and strain typing information are compiled in Table 1. Nine PVY isolates were collected from Idaho and Montana potato seed trials in 2011, 2012, and 2013, or from Othello, WA, seed lot trials in 2011 and 2012. Isolates Linda14 and Pondo4 were provided by Kerstin Lindner (Julius Kühn Institut, Braunschweig, Germany). Isolate T1 was provided by Dr. J. Whitworth (USDA-ARS, Aberdeen, ID). All isolates were maintained on tobacco (*Nicotiana tabacum* cv Burley) at the University of Idaho or at Cornell University, in insect-free growth rooms, with periodic mechanical re-inoculations.

4.2. Serological and RT-PCR strain typing

Prior to sequencing, each PVY isolate was typed to strain using serological profiling with three monoclonal antibodies distinguishing four serotypes previously identified for PVY strains PVY^{O}, PVY^{N}, PVY^{O}-O5, and PVY^{AST} (Karasev et al., 2010; Galvino-Costa et al., 2012a; Nikolaeva et al., 2012), and using RT-PCR with one or two differentiating primer sets (Lorenzen et al., 2006b; Chikh-Ali et al., 2013). Reverse transcription and subsequent PCR steps followed the protocols described previously (Hu et al., 2009a; Karasev et al., 2011). PCR products were separated on a 1% or 1.2% agarose gel and visualized after staining with GelStar (Lonza) or ethidium bromide.

4.3. Sequencing

Each isolate of PVY was propagated on tobacco and the fraction of total virus-specific nucleic acids was used for sequencing. Virus RNA extraction and cDNA synthesis were performed as described previously (Hu et al., 2009a; Karasev et al., 2011). Whole genomes of PVY isolates were sequenced directly from overlapping RT-PCR fragments amplified using a set of 48 near-universal PVY primer pairs developed for this study (Table 2) and the GreenTaq Taq-polymerase (Genscript, Piscataway, NJ) as described previously (Hu et al., 2009a; Karasev

Table 6
Summary of models selected by DT-Models for each of the five major sections.

Section #	Genome range (nt)	Size of section	Model chosen	Rate matrix
1	1–500	500	TVM+I+G	abcdbe
2	501–2390	1890	TrN+I+G	abaaea
3	2391–5850	3460	TrN+I+G	abaaea
4	5851–9200	3350	GTR+I+G	abcdef
5	9201–9704	505	HKY+I+G	abaaba

et al., 2011). This approach was found to be quite efficient and allowed us to sequence over 100 whole PVY genomes representing 9 different strains and recombinant types. Successfully amplified PCR products were treated with Exosap-It (Affymetrix, Cleveland, OH) and submitted for Sanger sequencing to Genewiz, Inc. (South Plainfield, NJ). Individual sequence reads were assembled using the SeqMan program of the Lasergene 9 Suite (DNASTAR).

4.4. Recombinant analysis

Recombinant analysis was performed on all whole PVY genomes using RDP4.22 in order to correctly identify the isolate strain types and cross-check them against the phylogenies as previously determined. Six of the available recombination analysis programs (RDP, GENECONV, Chimaera, MaxChi, Bootscan, and SiScan) were used with default settings to identify potential recombinants and parents, with an isolate determined to be a recombinant if all six had significant support ($p < 0.0001$) (Martin and Rybicki, 2000; Padidam et al., 1999; Posada and Crandall, 2001; Smith, 1992; Salminen et al., 1995; Gibbs et al., 2000).

4.5. Sequence sources and phylogenetic analysis

The 119 new sequences were combined with 166 from the GenBank database for further analysis (Tables 1, 3). Of the 166 whole PVY genomes extracted from GenBank, 19 represented strain PVY^{O}, 35 belonged to the PVY^{O}-O5 lineage, 33 represented strain PVY^{NTN} (28 PVY^{NTNa} and 5 PVY^{NTNb}), 3 were typed as PVY-NE11, 9 as $PVY^{N:O}$, 16 as $PVY^{N\text{-}Wi}$, 11 as $PVY^{NA\text{-}N}$, 4 as $PVY^{Eu\text{-}N}$, 5 as PVY^{C}, 2 as PVY^{E}, 12 belonged to three Syrian types (5 PVY-SYR-I, 6 PVY-SYR-II, and 1 PVY-SYR-III), and 19 were unclassified (Tables 3, 4).

Sequence alignment of the 285 genomes was conducted using either Clustal X or MUSCLE, with some manual adjustment (Larkin et al., 2007; Edgar, 2004). A whole-genome UPGMA tree was generated in RDP4.22 in order to quickly type each isolate to strain based on how they clustered together, and strain types were checked against RT-PCR and serological data, when available, and frequently also checked with BLAST (http://blast.ncbi.nlm.nih.gov/) (Sokal and Michener, 1958; Martin et al., 2010). All genomes were then divided into 5 major recombinant sections (as described in the Results) using Seqret, and model selection was run on each section separately using DTModSel (Rice et al., 2000; Minin et al., 2003). Models selected are listed in Table 6. Then, 100 maximum likelihood (ML) trees were constructed in Garli using the model information acquired from DTModSel, each run had 3,000,000 generations or until there were no longer improvements between generations (Zwickl, 2006). Bootstrap values were also calculated using Garli. The Figtree v1.4 program was then used to visualize and color the inferred trees (http://tree.bio.ed.ac.uk/software/figtree/). Strain diversity within and between all nonrecombinant strains was inferred from an identity matrix built using the BioEdit version 7.2.5 Sequence Identity Matrix function (Table 5) (Hall, 1999).

Acknowledgements

This project was funded in part through grants from the United States Department of Agriculture (USDA)-NIFA-NRI (2009-35600-05025), USDA-NIFA-SCRI (2009-51181-05894 and 2014-51181-22373), USDA-ARS (58-5354-7-540, 58-5354-2-345, and 58-1907-8-870), Idaho State Department of Agriculture, Idaho Potato Commission, and Washington State Potato Commission. The authors thank Teresa Meacham, Dr. Xiaojun Hu, and Zachary Sielaff for isolate typing and sequencing help at the initial stages of the project, as well as Dr. Yu-hsuan Lin and Jason Ingram for providing some field isolates and catching a few minor sequence errors.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.virol.2017.03.018.

References

Adams, M.J., Zerbini, F.M., French, R., Rabenstein, F., Stenger, D.C., Valkonen, J.P.T., 2012. Family *Potyviridae*. In: King, A., Adams, M., Carstens, E., Lefkowitz, E. (Eds.), In Virus Taxonomy. Ninth Report of the International Committee on Taxonomy of Viruseseditors. Elsevier, Oxford, 1069–1089.

Blanchard, A., Rolland, M., Lacroix, C., Kerlan, C., Jacquot, E., 2008. Potato virus Y: a century of evolution. Curr. Top. Virol. 7, 21–32.

Chikh-Ali, M., Bosque-Perez, N., Vander Pol, D., Sembel, D., Karasev, A.V., 2016a. Occurrence and molecular characterization of recombinant *Potato virus* Y^{NTN} (PVY^{NTN}) isolates from Sulawesi, Indonesia. Plant Dis. 100, 269–275.

Chikh-Ali, M., Alruwaili, H., Vander Pol, D., Karasev, A.V., 2016b. Molecular characterization of recombinant strains of *Potato virus Y* from Saudi Arabia. Plant Dis. 100, 292–297.

Chikh-Ali, M., Vander Pol, D., Nikolaeva, O.V., Melzer, M.J., Karasev, A.V., 2016c. Biological and molecular characterization of a tomato isolate of potato virus Y (PVY) of the PVY^{C} lineage. Arch. Virol. 161, 3561–3566.

Chikh-Ali, M., Rowley, J.S., Kuhl, J.C., Gray, S.M., Karasev, A.V., 2014. Evidence of a monogenic nature of the *Nz* gene conferring resistance against *Potato virus Y* strain Z (PVY^{Z}) in potato. Am. J. Potato Res. 91, 649–654.

Chikh-Ali, M., Gray, S.M., Karasev, A.V., 2013. An improved multiplex IC-RT-PCR assay distinguishes nine strains of *Potato virus Y*. Plant Dis. 97, 1370–1374.

Chikh Ali, M., Maoka, T., Natsuaki, T., Natsuaki, K.T., 2010. $PVY^{NTN\text{-}NW}$, a novel recombinant strain of *Potato virus Y* predominating in potato fields in Syria. Plant Pathol. 59, 31–41.

Chikh Ali, M., Maoka, T., Natsuaki, K.T., 2007. The occurrence and characterization of new recombinant isolates of PVY displaying shared properties of $PVY^{N}W$ and PVY^{NTN}. J. Phytopathol. 155, 409–415.

Chung, B.Y., Miller, W.A., Atkins, J.F., Firth, A.E., 2008. An overlapping essential gene in the Potyviridae. Proc. Natl. Acad. Sci. USA 105, 5897–5902.

Cockerham, G., 1970. Genetic studies on resistance to potato viruses X and Y. Heredity 25, 309–348.

Cuevas, J.M., Delaunay, A., Visser, J.C., Bellstedt, D.U., Jacquot, E., Elena, S.F., 2012. Phylogeography and molecular evolution of *Potato virus Y*. PLoS One 7, e37853. http://dx.doi.org/10.1371/journal.pone.0037853.

de Bokx, J.A., Huttinga, H., 1981. Potato Virus Y. Descriptions of Plant Viruses, No. 242. Commonw. Mycol. Inst./Assoc. Appl. Biol., Kew, England. Online ⟨www.dpvweb.net/dpv/showdpv.php?Dpvno=242⟩.

Djilani-Khouadja, F., Glais, L., Tribodet, M., Kerlan, C., Fakhfakh, H., 2010. Incidence of potato viruses and characterisation of *Potato virus Y* variability in late season planted potato crops in Northern Tunisia. Eur. J. Plant Pathol. 126, 479–488.

Dougherty, W.G., Carrington, J.C., 1988. Expression and function of potyviral gene products. Ann. Rev. Phytopathol. 26, 123–143.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, (1792-97).

Funke, C.N., Nikolaeva, O.V., Green, K.J., Tran, L.T., Chikh-Ali, M., Quintero-Ferrer, A., Cating, R.A., Frost, K.E., Hamm, P.B., Olsen, N., Pavek, M.J., Gray, S.M., Crosslin, J.M., Karasev, A.V., 2017. Strain specific resistance to *Potato virus Y* (PVY) in potato and its effect on the relative abundance of PVY strains in commercial potato fields. Plant Dis. 101, 20–28.

Galvino-Costa, S.B.F., Figueira, Ad.R., Camargos, V.V., Geraldino, P.S., Hu, X.-J., Nikolaeva, O.V., Kerlan, C., Karasev, A.V, 2012a. A novel type of *Potato virus Y* recombinant genome, determined for the genetic strain PVY^{E}. Plant Pathol. 61, 388–398.

Galvino-Costa, S.B.F., Figueira, A.R., Rabelo-Filho, F.A.C., Moraes, F.H.R., Nikolaeva, O.V., Karasev, A.V., 2012b. Molecular typing of *Potato virus Y* isolates from Brazil reveals a diverse set of recombinant strains. Plant Dis. 96, 1451–1458.

Gibbs, A., Ohshima, K., Yasaka, R., Mohammadi, M., Gibbs, M.J., Jones, R.A.C., 2017. The phylogenetics of the global population of potato virus Y and its necrogenic recombinants. Virus Evol.. http://dx.doi.org/10.1093/ve/vex002, 3, vex002.

Gibbs, A., Ohshima, K., 2010. Potyviruses and the digital revolution. Ann. Rev. Phytopathol. 48, 205–223.

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.

Glais, L., Tribodet, M., Kerlan, C., 2002. Genomic variability in Potato potyvirus Y (PVY): evidence that PVY^{N}W and PVY^{NTN} variants are single to multiple recombinants between PVY^{o} and PVY^{N} isolates. Arch. Virol. 147, 363–378.
Gray, S., De Boer, S., Lorenzen, J., Karasev, A., Whitworth, J., Nolte, P., Singh, R., Boucher, A., Xu, H., 2010. *Potato virus Y*: an evolving concern for potato crops in the United States and Canada. Plant Dis. 94, (1384–97).
Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
Hu, X., Meacham, T., Ewing, L., Gray, S.M., Karasev, A.V., 2009a. A novel recombinant strain of Potato virus Y suggests a new viral genetic determinant of vein necrosis in tobacco. Virus Res. 143, 68–76.
Hu, X., Karasev, A.V., Brown, C.J., Lorenzen, J.H., 2009b. Sequence characteristics of *potato* virus Y recombinants. J. Gen. Virol. 90, 3033–3041.
Jones, R.A.C., 1990. Strain group specific and virus specific hypersensitive reactions to infection with potyviruses in potato cultivars. Ann. Appl. Biol. 117, 93–105.
Karasev, A., Nikolaeva, O.V., Hu, X., Sielaff, Z., Whitworth, J., Lorenzen, J.H., Gray, S.M., 2010. Serological properties of ordinary and necrotic isolates of *Potato virus Y*: a case study of PVY^{N} misidentification. Am. J. Potato Res. 87, 1–9.
Karasev, A.V., Hu, X., Brown, C.J., Kerlan, C., Nikolaeva, O.V., Crosslin, J.M., Gray, S.M., 2011. Genetic diversity of the ordinary strain of *Potato virus Y* (PVY) and origin of recombinant PVY strains. Phytopathology 101, 778–785.
Karasev, A.V., Gray, S.M., 2013a. Continuous and emerging challenges of *Potato virus Y* in potato. Ann. Rev. Phytopathol. 51, 571–586.
Karasev, A.V., Gray, S.M., 2013b. Genetic diversity of *Potato virus Y* complex. Am. J. Potato Res. 90, 7–13.
Kehoe, M.A., Jones, R.A.C., 2016. Improving *Potato virus Y* strain nomenclature: lessons from comparing isolates obtained over a 73-year period. Plant Pathol. 65, 322–333.
Kerlan, C., 2006. Description of Plant Viruses: Potato virus Y. Association of Applied Biologists, No. 414 (⟨http://www.dpvweb.net/dpv/showdpv.php?Dpvno=414⟩).
Kerlan, C., Nikolaeva, O.V., Hu, X., Meacham, T., Gray, S.M., Karasev, A.V., 2011. Identification of the molecular make-up of the *Potato virus Y* strain PVY^{Z}: genetic typing of PVY^{Z}-NTN. Phytopathology 101, 1052–1060.
Kerlan, C., Tribodet, M., Glais, L., Guillet, M., 1999. Variability of *Potato virus Y* in potato crops in France. J. Phytopathol. 147, 643–651.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.
Lorenzen, J.H., Meacham, T., Berger, P.H., Shiel, P.J., Crosslin, J.M., Hamm, P.B., Kopp, H., 2006a. Whole genome characterization of *Potato virus Y* isolates collected in the western USA and their comparison to isolates from Europe and Canada. Arch. Virol. 151, 1055–1074.
Lorenzen, J.H., Piche, L.M., Gudmestad, N.C., Meacham, T., Shiel, P., 2006b. A multiplex PCR assay to characterize *Potato virus Y* isolates and identify strain mixtures. Plant Dis. 90, 935–940.
Lorenzen, J., Nolte, P., Martin, D., Pasche, J.S., Gudmestad, N.C., 2008. NE-11 represents a new strain variant class of *Potato virus Y*. Arch. Virol. 153, 517–525.
Martin, D.P., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.
Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.
Minin, V., Abdo, Z., Joyce, P., Sullivan, J., 2003. Performance-based selection of likelihood models for phylogeny estimation. Syst. Biol. 52, 674–683.
Moury, B., 2010. A new lineage sheds light on the evolutionary history of *Potato virus Y*. Mol. Plant Pathol. 11, 161–168.
Nagy, P.D., 2008. Recombination in plant viruses. In: Roossinck, M.J. (Ed.), Plant Virus Evolutionedited by. Springer Berlin Heidelberg, Berlin and Heidelberg, 133–156.
Nagy, P.D., Simon, A.E., 1997. New insights into the mechanisms of RNA recombination. Virology 235, 1–9.
Nie, B., Singh, M., Murphy, A., Sullivan, A., Xie, C., Nie, X., 2012. Response of potato cultivars to five isolates belonging to four strains of *Potato virus Y*. Plant Dis. 96, 1422–1429.
Nikolaeva, O.V., Roop, D., Galvino-Costa, S.F.B., Figueira, A.R., Gray, S.M., Karasev, A.V., 2012. Epitope mapping for monoclonal antibodies recognizing tuber necrotic strains of *Potato virus Y*. Am. J. Potato Res. 89, 121–128.
Ogawa, T., Tomitaka, Y., Nakagawa, A., Ohshima, K., 2008. Genetic structure of a population of *Potato virus Y* inducing potato tuber necrotic ringspot disease in Japan; comparison with North American and European populations. Virus Res. 131, 199–212.
Ogawa, T., Nakagawa, A., Hataya, T., Ohshima, K., 2012. The genetic structure of populations of *Potato virus Y* in Japan; based on the analysis of 20 full genomic sequences. J. Phytopathol. 160, 661–673.
Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–225.
Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. USA 98, 13757–13762.
Quenouille, J., Vassilakos, N., Moury, B., 2013. *Potato virus Y*: a major crop pathogen that has provided major insights into the evolution of viral pathogenicity. Mol. Plant Pathol. 14, 439–452.
Quintero-Ferrer, A., Robles-Hernandez, L., Gonzalez-Franco, A.C., Kerlan, C., Karasev, A.V., 2014. Molecular and biological characterization of a recombinant isolate of *Potato virus Y* from Mexico. Arch. Virol. 159, 1781–1785. http://dx.doi.org/10.1007/s00705-013-1968-0.
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16, 276–277.
Riechmann, J.L., Lain, S., Garcia, J.A., 1992. Highlights and prospects of potyvirus molecular biology. J. Gen. Virol. 73, 1–16.
Roossinck, M.J., 2003. Plant RNA virus evolution. Curr. Opin. Microbiol. 6, 406–409.
Salminen, M.O., Carr, J.K., Burke, D.S., McCutchan, F.E., 1995. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res. Hum. Retrovir. 11, 1423–1425.
Schubert, J., Thieme, T., Thieme, R., Ha, C.V., Hoang, G.T., 2015. Molecular and biological characterization of *Potato virus Y* isolates from Vietnam. J. Phytopathol. 163, 620–631.
Schubert, J., Fomitcheva, V., Sztangret-Wiśniewska, J., 2007. Differentiation of *Potato virus Y* strains using improved sets of diagnostic PCR-primers. J. Virol. Methods 140, 66–74.
Simon, A.E., Bujarski, J.J., 1994. RNA-RNA Recombination and evolution in virus-infected plants. Ann. Rev. Phytopathol. 32, 337–362.
Singh, R.P., Valkonen, J.P.T., Gray, S.M., Boonham, N., Jones, R.A.C., Kerlan, C., Schubert, J., 2008. Discussion paper: the naming of *Potato virus Y* strains infecting potato. Arch. Virol. 153, 1–13.
Smith, J.M., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.
Sokal, R., Michener, C., 1958. A statistical method for evaluating systematic relationships. Univ. Kans. Sci. Bull. 38, 1409–1438.
Visser, J.C., Bellstedt, D.U., Pirie, M.D., 2012. The recent recombinant evolution of a major crop pathogen *Potato Virus Y*. PLoS One 7, e50631. http://dx.doi.org/10.1371/journal.pone.0050631.
Wei, T., Zhang, C., Hong, J., Xiong, R., Kasschau, K.D., Zhou, X., Carrington, J.C., Wang, A., 2010. Formation of complexes at plasmodesmata for potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. PLoS Pathog. 6, e1000962.
Zwickl, D.J., 2006. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion (Ph.D. dissertation). The University of Texas at Austin, Austin, Texas.