POTATO VIRUS Y EVOLUTION AND RECOMBINATION

A Thesis

Presented in Partial Fulfillment of the Requirements for the

Degree of Master of Science

with a

Major in Bioinformatics and Computational Biology

in the

College of Graduate Studies

University of Idaho

by

Kelsie J Evans

May 2014

Major Professor: Alexander V. Karasev, Ph.D.

Authorization to Submit Thesis

This thesis of Kelsie J Evans, submitted for the degree of Master of Science with a major in Bioinformatics and Computational Biology and titled "*Potato Virus Y* Evolution and Recombination," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: _____ Date: _____
Alexander V. Karasev, Ph.D.

Committee Members: _____ Date: _____
Celeste J. Brown, Ph.D.

_____ Date: _____
James A. Foster, Ph.D.

_____ Date: _____
Christopher J. Williams, Ph.D.

Department Administrator: _____ Date: _____
Eva Top, Ph.D.

Discipline's College Dean: _____ Date: _____
Paul Joyce, Ph.D.

Final Approval and Acceptance by the College of Graduate Studies:

Dean of the College of Graduate Studies _____ Date: _____
Jie Chen, Ph.D.

Abstract

Potato virus Y (PVY) exists as a complex of strains, including a growing number of recombinants. Evolution of PVY proceeds rapidly through recombination, leading to adaptation of the virus to multiple potato cultivars and a wider range of environmental conditions. The origins of PVY recombinants were studied through whole genome sequencing of PVY genomes and subsequent phylogenetic and recombination analyses. A collection of 119 newly sequenced PVY isolates and 166 PVY whole genomes from the GenBank database was subjected to phylogenetic analysis focusing on genome sections commonly involved in recombination. Evolutionary relationships of these sections were thusly determined and a substantial diversity was revealed within both non-recombinant and recombinant strain types, with several lineages identified. Recombination analysis was then conducted to find novel recombinants, and 12 were identified. There are now 28 known recombinant patterns. It is now hypothesized that all known recombinant types of PVY originated more than once.

Acknowledgements

First I must extend tremendous gratitude to my advisor, Dr. Alexander Karasev, for constant and consistent direction, assistance, and patience. Also due many thanks is Dr. Celeste Brown, for extensive help with the more computational side of things, the ins-and-outs of the BCB program, and for serving on my committee. I would also like to thank Drs. James Foster and Christopher Williams for their guidance and support on my committee.

I am thankful for the advice, help, and friendship of current and previous members of the Karasev lab, especially Jenny Rowley, Arturo Quintero Ferrer, Dr. Mohamad Chikh Ali, Dr. Sahar Eid, Eunice Beaver-Kanuya, Cassandra Sago, Teresa Meacham, Dr. Olga Nikolaeva, Dr. Xue Feng, Mariana Rodriguez[2], Dr. Suellen B. F. Galvino-Costa, and Archana Shrestha.

I am fortunate to have previously worked alongside or collaborated with several wonderful and helpful people to whom I must extend many thanks, particularly Valorie Wiss, Dr. Pavitra Roychoudhury, Amber Stancik, Luyi Zhao, Dr. Stewart Gray, and Dr. Yu-hsuan Lin. I am also grateful for the assistance of Rob Lyon and Benji Oswald when it came to using the IBEST computers, to Dr. Jack Sullivan and Brice Sarver for guidance with choosing and using certain programs, and to Dr. Xiaojun Hu for inspiring the skeleton of my project.

I would also like to thank my family for their support and love, especially my husband, Logan Evans, who offered me unwavering support and encouragement and also was able to help me with computer stuff on top of everything else.

Last but certainly not least, I am deeply grateful for all of my friends and their unending encouragement, humor, nerdiness, and love, especially Vandhana Krishnan, Shiann Johns, and my BFFs Karina Hayden and Leah Girodat.

Dedication

This thesis is dedicated to my fellow nonbelievers. Stay strong in the many faces of adversity, and know that science is always more powerful than superstition.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: General Introduction

Introduction

Background

Potato virus Y (PVY) is the most economically important virus damaging potato crops worldwide, affecting both tuber quality and yield and costing millions of dollars annually (Gray et al. 2010; Karasev and Gray, 2013). PVY is the type member of the family *Potyviridae* and has a single-stranded RNA genome ca. 9.7-kb in length, with a flexuous, filamentous particle structure (Berger et al., 2005). The PVY genome encodes one long polyprotein which is post-translationally cleaved into ten mature proteins by three virus-specific proteolytic enzymes (Fauquet et al., 2005; Dougherty and Carrington, 1988). An eleventh, frame-shifted protein called PIPO has also been recently discovered (Chung et al., 2008). The 5' terminus of PVY has a covalently-linked VPg protein, and the 3' terminus has a poly(A) tail (Riechmann et al., 1992). PVY affects primarily potato but can also infect tomato, pepper, tobacco, and tomato crops (McDonald and Singh, 1996). However, the focus here will be almost entirely on potato.

Spread

PVY is an aphid-transmitted virus, although it can also be transmitted mechanically with some difficulty, and vegetatively through potato seed. However, PVY cannot be transmitted through true seed; from a practical perspective, transmission through potato seed and by more than fifty species of aphids are the most important (Kerlan, 2006). The reasons for the wide spread of PVY in potato crops world-wide, including the U.S., are not completely understood and may be related to several factors that include emergence of new recombinants, influx of new cultivars of potato, production of potato in a multitude of new environments, or

a combination of these and other yet unknown factors. PVY causes two types of disease symptoms in potato, a) foliage symptoms, such as mosaic, crinkling, stunting, and a range of local and systemic necrotic reactions; and b) tuber symptoms, from a simple yield reduction to the most important: potato tuber necrotic ringspot disease (PTNRD), which renders tubers unmarketable and may result in up to a 100% crop loss (Kerlan, 2006; Gray et al., 2010; Karasev and Gray, 2013). The exact virus genetic determinants for each of these symptom are currently unknown, but symptom expression broadly depends on the cultivar of potato, the strain of PVY, and environmental factors.

PVY Recombination

The large range of symptoms expressed by PVY in potato is due to the fact that PVY exists as a complex of strains. These strains can be distinguished based on serology, reaction on indicator cultivars, and the recombinant structure of their genomes. Strains comprise non-recombinant (e.g. PVY^{O}, PVY^{C}, PVY^{O5}, $PVY^{NA\text{-}N}$, and $PVY^{Eu\text{-}N}$) and recombinant (e.g. $PVY^{N:O}$, $PVY^{N\text{-}Wi}$, PVY^{NTNa}, PVY^{NTNb}, $PVY^{NE\text{-}11}$, PVY^{E}) types (e.g. Karasev and Gray, 2013). Recombinants typically contain large (≥500nt) segments of different parental genomes (Hu et al., 2009). There is a general perception that both the number and diversity of recombinants in the field has vastly increased over the past five years, and it has been established that recombinant strains of PVY are typically associated with PTNRD (Gray et al., 2010; Karasev and Gray, 2013). These recombinants may have some selective advantages over non-recombinant strains, including milder symptoms (as compared to PVY^{O} or PVY^{O5} which can cause whole plant necrosis or death in certain cultivars), ability to evade detection by molecular methods during potato certification screening, and potentially the ability to survive in a wider variety of potato cultivars and environmental conditions.

Materials and Methods

Isolate Sources, Sequencing, and Genome Sources

In this thesis, a large collection of 285 whole-genome PVY isolates was used for phylogenetic and recombinant analysis. Of this set, 166 whole-genome sequences were extracted from the GenBank database. The remaining 119 isolates of PVY were sequenced by us, and represent field isolates largely collected from the U.S. potato. These were sequenced using a primer set developed for fast, comprehensive, and easy PVY Sanger sequencing (Table 1.1).

Isolate and Sequence Typing

Field isolates of PVY were typed initially by triple-antibody sandwich (TAS)-ELISA (able to differentiate four serotypes of PVY) and by two different multiplex reverse transcription (RT)-PCR assays (able to differentiate at least nine strains of PVY). After sequencing, these isolates were added to the GenBank sequences and a master alignment of all 285 whole-genome PVY sequences was created using the MUSCLE program, with some manual adjustment (Edgar, 2004). A quick UPGMA tree was created for this alignment using RDP4.22 and isolates were typed to strain based on where they clustered (Martin et al., 2010). Later, isolates were also subjected to recombination analysis when it was discovered that UPGMA was not sufficient to catch novel recombinants. However, it was adequate approximately 95.4% of the time (only 13 whole genome sequences out of 285 were typed incorrectly via UPGMA).

Phylogenetic Analysis

This master alignment was then divided into the five common recombinant sections (with breakpoints at nucleotide positions 500, 2390, 5850, and 9200) and a model was

selected for each section by the DT-ModSel program (Minin et al., 2003). Using these models, the Garli program was then run on each section for 100 replicates and a bootstrapped phylogeny was produced using the Sumtrees program (Zwickl 2006; Sukumaran and Holder, 2010).

Recombination Analysis

This same set of isolates was then subjected to recombination analysis using six programs in the RDP4.22 program suite (Martin et al., 2010). Whole-genome sequences of isolates were analyzed in small groups according to their previous UPGMA typing, with at least one representative isolate of each potential parental strain included.

Results

Sequencing the 119 Novel PVY Whole Genomes

A set of 48 primer pairs, suitable for rapid amplification and sequencing of a variety of PVY strains, was designed. Then, 119 new genomes were sequenced in the space of 24 months, belonging to eight different strains and recombinant structures, including PVY^{C} which was never before sequenced in the U.S.

Phylogenetic Analysis

Based on clade placement for each of the five major sections, PVY strains were separated into different sub-groups within each strain. For the non-recombinant parents, there are five distinct clades of O, three of NA-N, and two of Eu-N. For the recombinants, there are three distinct clades of NTNa, two of NTNb, three of N-Wi, only one of N:O (other than isolate MI090004 which is separate for sections 1 and 2, and isolate ND23 which is arguably not N:O at all), one of E, and two of NE-11.

Recombination Analysis

From recombination analysis, twelve novel, previously unreported PVY recombinant patterns were found. Three were from field isolates of U.S. potato (AL100001, ND23, and NY110001). The other nine came from GenBank. The O5 strain has never been found to serve as a parent, and conversely, the non-recombinant parent of the largest NE-11 section has not yet been found.

Conclusions and Discussion

PVY has far greater diversity than previously thought, in terms of diversity of non-recombinant potential parent strains, in terms of common recombinant strains, and in terms of recombinant structures. Current typical strain typing methods are not sufficient to type a PVY isolate to strain correctly every time. The diversity of recombinants has arisen from diversity in the parents. This study also underlines the importance of using proper tools to check the recombinant pattern of each sequenced isolate. It is easy to miss novel recombinants.

References

Berger, P.H., Adams, M.J., Barnett, O.W., Brunt, A.A., Hammond, J., Hill, J.H., Jordan, R.L., Kashiwazaki, S., Rybicki, E., Spence, N., Stenger, D.C., Ohki, S.T., Uyeda, I., van Zaayen, A., Valkonen, J., Vetten, H.J., (2005). Family Potyviridae. Virus taxonomy, Classification, Nomenclature of Viruses. Eighth Report of the International Committee on Taxonomy of Viruses. Academic Press, pp. 819–841.

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.

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 Research 32(5), 1792-97.

Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A., (2005). (Eds.), Virus Taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA, USA.

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 Disease 94, 1384–97.

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.

Karasev, A.V. and Gray, S.M., (2013) Continuous and emerging challenges of Potato virus Y in potato. Annual Review of Phytopathology 51: 571-586.

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).

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(5), 674-683.

McDonald, J.G. and Singh, R.P., (1996). Host range, symptomology and serology of isolates of Potato virus Y (PVY) that share properties with both the PVY^{N} and PVY^{O} strain groups. Amer. Pot. J., 73: 309- 314.

Riechmann, J.L., Lain, S., Garcia, J.A., (1992). Highlights and prospects of potyvirus molecular biology. J. Gen. Virol. 73, 1–16.

Sukumaran, J., Holder, M.T., (2010). A Python library for phylogenetic computing. Bioinformatics 26, 1569-1571.

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.

Supplementary Information

Table 1.1 Primer set developed and used for whole-genome PVY Sanger sequencing.

O-Type Pairs for Full Genome Sequencing

Pair #	Forward Primer	Forward Primer Sequence	Reverse Primer	Reverse Primer Sequence	Segment Length	Genome Span
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

Pair #	Forward Primer	Forward Primer Sequence	Reverse Primer	Reverse Primer Sequence	Segment Length	Genome Span
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

Pair #	Forward Primer	Forward Primer Sequence	Reverse Primer	Reverse Primer Sequence	Segment Length	Genome Span
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

CHAPTER 2: Phylogenetic Study of the Origins of Recombinant Strains of *Potato virus Y*

Kelsie J Evans[1,3], Celeste J. Brown[2,3], Stewart M. Gray[4], and Alexander V. Karasev[1,3]

[1]Department of PSES, [2]Department of Biological Sciences, and [3]Bioinformatics and Computational Biology Program, University of Idaho, Moscow, ID; [4]USDA-ARS, Cornell University, Ithaca, NY

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, combining large sections of parental genomes, which leads to adaptation of the virus to multiple potato cultivars and a wider range of environmental conditions. The role of recombination in PVY evolution and origin of the common PVY recombinants, such as $PVY^{N:O}$, PVY^{N-Wi}, and PVY^{NTN}, were studied here through whole genome sequencing of PVY genomes and subsequent phylogenetic and recombination analysis. A collection of 119 newly sequenced PVY isolates and 166 whole PVY genomes from the GenBank database was subjected to phylogenetic analysis, focusing on large genome sections commonly involved in recombination. Two new PVY^{C} recombinants were sequenced and identified. A substantial diversity was revealed within non-recombinant parental strain types PVY^{O} and PVY^{N}, with several distinct lineages. This diversity in the parental sequences allowed us to trace the origins and evolution of all recombinant types of PVY, which also showed considerable diversity in most cases. The data obtained suggest that all common

recombinant types originated more than once, from different parental sequences, and hence likely to have some selective fitness advantages.

Introduction

PVY is the type member of the genus *Potyvirus*, family *Potyviridae* and has a single-stranded RNA genome ca. 9.7-kb nucleotides in length with a flexible, filamentous structure (Berger et al., 2005). PVY has a single-stranded, positive-sense genome which encodes a single polyprotein later cleaved, co- and post-translationally, into ten mature proteins by three virus-specific proteases (Fauquet et al., 2005; 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 is thought to interact with protein P3 and assist with movement of the virus *in planta* (Chung et al., 2008; e.g. 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 exists as a complex of strains distinguished based on hypersensitive resistance (HR) response towards three *N* genes known in potato, *Ny*, *Nc*, and *Nz*. Isolates of PVY eliciting HR in the presence of the Ny_{tbr} gene 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; deBokx and Huttinga, 1981; Jones, 1990; Singh et al., 2008; Karasev and Gray, 2013; Chikh-Ali et al., 2014). Strains PVY^{N} and PVY^{E} comprise isolates of PVY that are unable to elicit HR in the presence of any of the three *N* genes, but PVY^{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.,

2012). Molecular genetics studies revealed that PVY isolates often have recombinant genomes composed of large segments from different parental genomes, PVY^{O} and PVY^{N} in most cases (Glais et al., 2002; Lorenzen et al., 2006; Hu et al., 2009). There are sixteen types of PVY recombinant structure reported to date, which include nine relatively common recombinants found in many geographical locations, namely $PVY^{N:O}$, PVY^{N-Wi}, PVY^{NTNa}, PVY^{NTNb}, PVY^{NE11}, PVY^{E}, and $PVY^{SYRI,II,\text{ and }III}$ (Lorenzen et al., 2006, 2008; Hu et al., 2009; Chikh Ali et al., 2007, 2010; Galvino-Costa et al., 2012a,b; Karasev and Gray, 2013), and seven rare recombinant types found and reported only once or twice, namely PVY^{N-Wi}-156var, PVY^{N-Wi}-261-4, PVY-SCRI-N, PVY-FrN, PVY-Nicola, PVY-T13, and PVY-nnp (e.g. Lorenzen et al, 2006; Schubert et al., 2007; Lorenzen et al., 2008; Chikh Ali et al., 2010; Galvino-Costa et al., 2012a; Ogawa et al., 2008, 2012; Karasev and Gray, 2013). Genomes of PVY^{O}, PVY^{N}, and PVY^{C} strains were found to be non-recombinant, serving as parents for many recombinant structures (Glais et al., 2002; Lorenzen et al., 2006). Conversely, only a few recombinants/recombinant types were subjected to a thorough biological characterization on potato indicators carrying different *N* genes, and hence the PVY^{NTN} recombinant was found to belong to the PVY^{Z} strain, while PVY^{E} was found to have a sophisticated recombinant structure with PVY^{NTN} and PVY^{NE11} serving as parents (Kerlan et al., 2011; Galvino-Costa et al., 2012a; Quintero-Ferrer et al., 2014). Considerable diversity was found in sequences of PVY isolates from the PVY^{O} strain, producing a distinct sub-lineage called PVY^{O}-O5 (or PVY^{O5}) which could also be distinguished biologically (Karasev et al., 2011; Nie et al., 2011).

Initial computational analysis of PVY whole genomes suggested that the number of recombinant patterns reported for PVY isolates from potato was relatively limited, and

positions of the main recombinant junctions (RJs) remarkably conserved (Hu et al., 2009). The driving forces for the emergence and survival of the PVY recombinants had not been yet elucidated, although there are no special sequences or secondary structures associated with the RJs (Hu et al., 2009). 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). Phylogenetic studies of PVY recombinants have been challenging because of the limited number of whole genomes used (Karasev et al., 2011; Ogawa et al., 2012; Visser et al., 2012; Quenouille et al., 2013), difficulties of accounting for recombination in building trees, ensuing necessities to analyze genome segments in order to reconstruct 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).

Here, we describe a large-scale sequencing project focused on a set of characterized potato isolates of PVY. Whole genomes were sequenced for 119 isolates of PVY representing seven different strains and recombinant types. These 119 whole genome sequences were subjected to a complete phylogenetic and recombination analysis together with 166 whole genome PVY sequences extracted from the GenBank database. In this study, common large recombinant sections were individually analyzed for this collection of 285 whole-genome sequences of PVY isolates in order to determine the relationships and origins of commonly recombinant sections of genomes.

Materials and Methods

Virus Sources, RNA Extraction and RT-PCR Amplification

Of the 119 newly sequenced PVY isolates, 107 came from a national PVY survey conducted in the United States between 2004-2006 (Gray et al., 2010). Nine PVY isolates were collected from Idaho potato seed winter grow-outs 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. Viral RNA was extracted from fresh tobacco tissue using the Dellaporta method as described previously (Hu et al., 2009a; Karasev et al., 2011). Prior to the sequencing step, each PVY isolate was typed to strain using serological profiling (Karasev et al., 2010; Nikolaeva et al., 2012) and RT-PCR with one or two differentiating primer sets (Lorenzen et al., 2006b; Chikh-Ali et al., 2013). PCR products were separated on a 1% or 1.2% agarose gel and visualized after staining with GelStar (Lonza). Reverse transcription and subsequent PCR steps followed the protocols described previously (Hu et al., 2009b; Karasev et al., 2011).

Sequencing

To sequence the whole genome of a PVY isolate, it was amplified by RT-PCR in approximately 40 overlapping DNA fragments ca. 1,000-bp long, using a set of PVY-specific primers designed specifically for this purpose (Table 1.1). Three major sets of primers were used for sequencing: a set designed for N-types, e.g. Eu-N, NA-N, NE-11, or sections of recombinants with these types as parents, a set designed for O-types, e.g. O, O5, or sections of

recombinants with these types as parents, and a collection of primers not intended to be strain specific, although we noted that even O- and N-type primers often turned out not to be strain-specific. 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). Whole genome sequences for all newly sequenced PVY isolates were deposited in the GenBank database, and corresponding accession numbers are listed in Table 2.1.

Sequence Sources and Phylogenetic Analysis

In total, 285 whole-genome PVY sequences were used for analysis (Table 2.1). Of these, 166 were from the GenBank database. The remaining 119 whole PVY genomes were newly sequenced at the University of Idaho; the specifics are described in the Results and in Table 2.1.

Sequence alignment 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 the isolate to strain based on how they clustered, 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/Blast.cgi) (Sokal and Michener, 1958; Martin et al., 2010). Genomes were then all divided into the 5 major recombinant sections as described in the Results section using Seqret, and model selection was run on each section separately using DTModSel (Rice et al., 2000; Minin et al., 2003). Models selected are in Table 2.2. Then, 100 bootstrapped maximum likelihood (ML) trees were constructed in Garli using the model information acquired in DTModSel, each time

run for 3,000,000 generations or until there were no longer improvements between generations (Zwickl, 2006). The very best tree (best ML score) of all the runs was selected as the reference tree and the Sumtrees program was used to create bootstraps on this reference tree based on all 100 of the best trees (Sukumaran and Holder, 2010). The MEGA5 program suite was used to create a condensed tree which collapsed all nodes with less than 70% bootstrap support (Tamura et al., 2011). The resulting trees are Figs. 2.2-2.6.

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 (Figs. 2.2-2.6). Six of the available recombination analysis programs (RDP, GENECONV, Chimaera, MaxChi, Bootscan, and SiScan) were used to identify potential recombinants and parents (Martin and Rybicki, 2000; Padidam et al., 1999; Posada and Crandall, 2001; Smith, 1992; Salminem et al., 1995; Gibbs et al., 2000).

Results and Conclusions

Sequencing 119 Whole Genomes of PVY Isolates

To sequence multiple whole genomes of PVY that belonged to different strains and had different recombinant structures, a set of 48 primer pairs was developed (Table 1.1) which allowed us to amplify the entire genome for any PVY isolate tested. This set was divided into O-primers (primer pairs 1 to 16) predominantly amplifying O-sequences, N-primers (primer pairs 17 to 32) predominantly amplifying N-sequences, and miscellaneous primers (primer pairs 33 to 48) used to amplify sequences other than O and N types specifically. In this approach, an isolate of PVY was subjected to an initial typing via serological profiling and RT-PCR (see Materials and Methods), to get a sense of what

parental sequences may be present in its genome, then cDNA was produced from the genomic RNA using random priming, and a set of overlapping PCR fragments was produced for each PVY isolate using RT-PCR with all primer pairs (see Table 1.1), and these overlapping PCR fragments were subjected to sequencing from both ends. The primer set was developed to produce ca. 1,000-bp PCR fragments that overlapped each other for about 500-bp. In this case, sequencing from both termini produced sufficient overlap in the middle of the 1-kb PCR fragment. Consequently, 34-40 primer pairs were sufficient to amplify the entire 9.7-kb genome of PVY at least two times, for a minimum of 4-fold sequence coverage. This approach was found quite efficient and allowed us to sequence over 100 whole PVY genomes representing 9 different strains and recombinant types.

The GenBank accession numbers for all sequenced PVY whole genomes are listed in the Supplementary Table 2.1. Of the 119 whole PVY genomes determined in this work, 26 represented strain PVY^{O}, 24 belonged to the PVY^{O}-O5 lineage, 15 represented strain PVY^{NTN}, 8 were typed as PVY-NE11, 23 as $PVY^{N:O}$, 17 as PVY^{N-Wi}, 4 as PVY^{N}, and 3 were not classified prior to sequencing, although serology and RT-PCR typing indicated possible O-types.

Dataset for the Whole PVY Genome Analysis

In addition to the 119 whole PVY genomes determined in this work, 166 whole genome sequences were downloaded from the GenBank database (see Table 2.1) to further diversify the dataset for phylogenetic analysis. 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 NTNa and 5 NTNb), 3 were typed as PVY-NE11, 9 as $PVY^{N:O}$, 16 as PVY^{N-Wi}, 11 as PVY^{NA-N}, 4 as PVY^{N}, 5 as PVY^{C}, 2 as PVY^{E}, 12 belonged to

three Syrian types (five SYR-I, six SYR-II, and one SYR-III), and 19 were unclassified. There was a slight bias towards 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.

Recombinant Analysis

Nearly 96% of all PVY isolates studied (272/285) were firmly typed based on serology, RT-PCR data, and a quick whole-genome UPGMA tree alone. This is probably due to a relatively limited set of main recombinants characteristic of PVY (Hu et al., 2009a; Karasev and Gray, 2013). Figure 2.1 summarizes the recombinant structures revealed in the entire PVY dataset here after the RDP analysis. Five parental, non-recombinant genomes were assigned to PVY^{O}, PVY^{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. 2.1), often in multiple recombinant types. Nevertheless, the majority of the recombinants were found to have PVY^{O} and PVY^{N} as parents. One additional parental sequence type constituted about 75% of the genome for the PVY-NE11 genotype, and also could be found in at least two different PVY recombinants (Fig. 2.7). Only one distinct sequence type, PVY^{O}-O5 could not be found in any of the recombinant genomes analyzed (Fig. 2.7). Seven unusual, or unique recombinant structures found previously in a very limited number of PVY whole genomes are highlighted on Fig. 2.7, along with three new ones determined in this study.

Two PVY isolates, NY110001 and AL100001, were identified as recombinants between PVY^{C} and PVY^{O} or $PVY^{NE\text{-}11}$ genomes, respectively, with novel recombinant structures (Fig. 2.7). In this case, a relatively small fragment of the PVY^{O} or $PVY^{NE\text{-}11}$

genome, respectively, could be seen in the otherwise PVY^{C} genome background inserted in the CI cistron. Both are laboratory isolates which were maintained in tobacco at the Cornell University laboratory. They displayed a typical O-serotype, binding O-specific monoclonal antibodies, and not binding N-specific monoclonal antibodies (not shown). When subjected to the RT-PCR typing using the protocol of Lorenzen et al. (2006b), both produced a single 267-bp band indicating a possible PVY^{O} isolate, but when subjected to the RT-PCR typing using the protocol of Chikh-Ali et al. (2013), they produced no bands, indicating a new, unknown genotype (not shown). Additionally, isolate ND23 was found to have a structure similar to that of a typical $PVY^{N:O}$, but with the RJ shifted left (from nt 500 to nt 383).

PVY Genome Sectioning and Phylogenetic Analysis

Due to the limited number of the RJs in the most common recombinants of PVY, and their relatively conserved positions (see Fig. 2.1), we attempted 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 come 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^{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 2.1). Phylogenies for sections 2-4 provided better resolution because of their larger size (Figs. 2.3-5), while trees for sections 1 and 5 were less robust and informative due to being shorter (Figs. 2.2 and 2.6). Overall, each of the sections 1-5 allowed clear separation in the phylogenetic trees between PVY^{O} and PVY^{C} types on one hand, and PVY^{N} and PVY^{NA-N} on the other hand.

The section 2 area represented almost exclusively non-recombinant sequences in the O-type genomes, and thus gave the best picture of the PVY^{O} diversity between genome positions 501 and 2,390 (Fig. 2.3), for all non-recombinant PVY^{O} genomes. At least five O-specific clades can be seen within the PVY^{O} lineage, with five additional clades, numbered O5-1 to O5-5 that included PVY^{O}-O5 sequences. Three of the identified O-specific clades corresponded to the clades PVY^{O}-1, (isolate ME120), PVY^{O}-2 (isolate CW), and PVY^{O}-3 (isolate Oz) delineated by us previously (Karasev et al., 2011). The two additional O-clades were designated PVY^{O}-4 (isolate NE38) and PVY^{O}-5 (isolate WI3) here, as marked on Fig. 2.3. The only two recombinant isolates, N-Egypt and PVY-Fr, present on this tree for the genome section 2, grouped very clearly with isolates of the PVY^{O}-4 clade (Fig. 2.3), suggesting that the O-specific parent for these two recombinants, N-Egypt and PVY-Fr, came from this clade. The separate O5 lineage itself demonstrated a substantial diversity in section 2. There were at least five distinct clades within O5, labeled 1 to 5, including two clades of O5- isolates missing the O5 serological marker, i.e. O5-4 (isolate WY1) and O5-5 (isolate O-139). Isolate T1 grouped closer to the O5 isolates but stood alone among isolates of the O5 clade.

The same section 2, when analyzed for N-type sequences, included predominantly sequences from various recombinant PVY genomes and demonstrated substantial diversity in this region. Aside from the NA-N lineage that clustered separately, other N-sequences demonstrated too high of a diversity that generally did not allow grouping of these sequences into meaningful clades. Only two possible exceptions could be found – a distinct clade combining thirty of the 32 N:O recombinants and a distinct clade with the two PVY^{E} recombinants (Fig. 2.3). However, one N:O recombinant, isolate MI090004, was clearly

grouped with a different recombinant type, SYR_II_2_8, while isolate ND23 was grouped with yet another recombinant type, N-Wi, thus suggesting at least 3 independent recombinant events resulting in an N:O recombinant pattern (Fig. 2.2). There are two distinct clades within the NE-11 lineage in Fig. 2.2. This split can be easily explained because of two slightly different positions (2220nt vs. 2009nt) of the RJ separating the 5'-terminal N-sequence and 3'-proximal NA-N sequence in isolates typified by ID20 on one hand, and by NE-11 on the other hand.

All fourteen NA-N sequences were grouped in a single clade but no other strain types formed separate monophyletic clades for this N-specific section 2 region. PVY^N sequences were present in at least two clades, NTNb in at least three, and NTNa in at least five clades (Fig. 2.2). The N-Wi recombinants did not form distinct clades in this section 2 phylogenies, and furthermore did not have any discernible relationships with other genome types besides MAF-VOY (N-Wi type) grouping with isolate ND23 (N:O type) for their section 2 (Fig. 2.2). Despite similarities in their genome structure, NTNa and NTNb isolates were not found phylogenetically related to one another for this section 2, nor were N:O and N-Wi isolates. Even within a specific recombination pattern, such as SYR-II, there was an apparent diversity of clade placement for different isolates. Some SYR-II isolates share a parent of their section 2 with SYR-I (such as 1108), some with unclassified strains (such as Wilga156), and some with none of the isolates included in these analyses, such as Be1 (Fig. 2.2). All these observations supported the conclusion that these recombinant structures could have arisen more than once from different parents providing specific fragments for their section 2.

Most of the sequences analyzed for section 3 represented O-type sequences from both recombinant and non-recombinant PVY genomes (see Fig. 2.1). O-phylogenies generated

from this section 3 allowed us to separate the same five clades for non-recombinant PVY^O isolates (Fig. 2.4) as section 2 phylogenies (Fig. 2.3), designated the same way as PVY^O-1 to -5. Recombinant PVY isolates carrying O-sequences in this section 3, including NTNa, NTNb, N:O, N-Wi, and SYR I to III types, grouped separately from the non-recombinant sequences analyzed (Fig. 2.4), although some recombinants, like N-Wi, NTNb, and some SYR I and II types were found related to the non-recombinant PVY^O-2 and PVY^O-3 clades (Fig. 2.4) and formed a larger clade which included non-recombinant isolates from PVY^O-2 and PVY^O-3 clades and also some recombinant isolates of N-Wi, NTNb, 156var, SYR I, and SYR II types (Fig. 2.4). NTNa isolates were all found in a single lineage comprising only recombinant types, NTNa, NTNb, SYR I, SYR II, and SYR III (Fig. 2.4). In this large recombinant lineage, NTNa isolates formed six separate clades in this section 3 phylogeny (Fig 2.4) rather than five found for section 2 (Fig. 2.3), although four of the clades from section 2 remained identical. Interestingly, none of the non-recombinant PVY^O isolates analyzed were related to the O parent providing the NTNa section 3 (see Fig. 2.4), hence no non-recombinant O-sequences were present in the recombinant lineage with NTNa isolates. However, some of the Syrian and NTNb isolates have a section 3 parent closer to the parent of the NTNa section 3 and some have a parent closer to the non-recombinant O isolates. Both PVY^E isolates were found in the same lineage with NTNa isolates and hence likely shared a parent with NTNa for section 3. Thirty-two N:O isolates were found monophyletic for this section 3, forming a distinct, single clade (Fig. 2.4), and no non-recombinant PVY^O isolates were found closely related to the hypothetical PVY^O parent of the N:O section 3. N-Wi isolates could be found in three clades, one of which included both recombinant and non-recombinant O-sequences (Fig. 2.4). Surprisingly, no O5 sequences were found in any of the recombinant PVY isolates,

although O5 sequences could be separated into at least six clades identical to the ones distinguished in the section 2 phylogeny (Fig. 2.3).

The number of N-type sequences analyzed for section 3 was limited, and all main N types were clearly distinguished in separate clades, N, NA-N, and NE-11 (Fig. 2.4). NE-11 sequences were separated into the same two clades (Fig. 2.4) as for section 2 (Fig. 2.3). 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. 2.4). PVY^{N} isolates did not form distinct clades as for section 2, but $PVY^{NA\text{-}N}$ isolates fall into the exact same two clades as for section 2 (Fig. 2.4, Fig. 2.3).

The O-sequences of section 4 (see Fig. 2.1) were represented by non-recombinant PVY^{O} isolates and also certain recombinant types, like N:O and N-Wi. The same five clades of non-recombinant PVY^{O}, designated PVY^{O}-1 to -5, could be found (Fig. 2.5) as for sections 2 and 3, with some sequences from the PVY^{O}-2 clade placed in a larger lineage, this time including both N:O and N-Wi recombinants. The three other non-recombinant PVY^{O} clades, PVY^{O}-1, PVY^{O}-4, and PVY^{O}-5, did not include recombinant PVY isolates (Fig. 2.5). N-Wi type recombinants were found in at least three clades (Fig. 2.5). For section 4, the N:O recombinant type was no longer found in its own distinct clade, but was rather included in a larger lineage combining recombinant and non-recombinant PVY^{O} isolates from the PVY^{O}-2 clade. The O5 sequences were still grouped in a single large lineage with at least six major clades, and no recombinants were found close to O5 sequences.

The N-sequences of section 4 were represented by non-recombinant PVY^{N}, $PVY^{NA\text{-}N}$, and also by various recombinants including NE-11, NTNa, NTNb, and SYR-I to –III (see Fig. 2.1). The non-recombinant NA-N isolates formed three of the same clades as for sections 2

and 3 (Fig. 2.5). The NE-11 isolates grouped into the same two clades as for sections 2 and 3, and now both PVY^{E} isolates grouped with only one of the NE-11 clades (with isolate ID20, Fig. 2.5). The NTNa isolates formed the same clades as for section 3 (Fig. 2.5). All NTNb isolates formed a single distinct clade (Fig. 2.5). The non-recombinant PVY^{N} formed two distinct clades exemplified by isolates Mont and N-605 (Fig. 2.5). Interestingly, all SYR isolates were placed separately from the main lineage combining all N, NA-N, NE-11, and NTN clades.

Discussion

Recombination, reassortment, and accumulation of mutations are the main forces shaping 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 of evolution for potyviruses (Gibbs and Ohshima, 2010). For PVY, the occurrence of multiple recombinant structures is well established (Glais et al., 2002; Lorenzen et al., 2006; Ogawa et al., 2008, 2012; Hu et al., 2009a,b; Karasev and Gray, 2013). Currently, recombinant strains of PVY dominate among PVY isolates circulating in potato in Europe, Africa, North America, and Brazil (Blanchard et al., 2008; Djilani-Khouadja et al., 2010; Gray et al., 2010; Galvino-Costa et al., 2012b; Visser et al., 2012). However, because PVY can be easily spread with potato seed over long distances, and extensive world trade in seed potato, it is difficult to trace the location origins of different recombinant types and reconstruct their evolution.

Since the number of PVY recombinants circulating in nature was found relatively limited (Hu et al., 2009b), the question was posed: how often do the recombinations between different PVY strains actually happen? 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). 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; Quenoille et al., 2013). One of the ways to avoid this is to exclude all recombinants from the analysis (Moury, 2010; Cuevas et al., 2012), but in this case the whole point of the analysis will be lost. Another way to deal with this recombination problem would be 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, like $PVY^{N:O}$ recombinants found to comprise respective isolates in a single, perhaps monophyletic clade, 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, we extended our phylogenetic analysis to all five individual sections of the PVY genome between the four main recombinant junctions (see Fig. 2.1), and substantially expanded the number of the PVY genomes analyzed, which included 119 newly sequenced genomes and 166 whole genomes from GenBank. These sequences represented all types of PVY recombinants and can be considered an unbiased set of PVY sequences suitable for a global analysis of PVY recombinants. Three questions that we tried to address were the same posed by us 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 recombinant events happen between PVY strains?

Examination of the phylogenetic trees presented in Figs 2.2-4 suggested that there might be no monophyletic lineages of PVY recombinants. Even the N:O lineage comprising 30 or 31 corresponding sequences out of 32 was not monophyletic, with 2 (Fig. 2.3) or 1 (Fig. 2.4) additional clades comprising MI090004 and ND23 isolates. Nevertheless, the ND23 isolate was found unusual, having the RJ2 shifted in the 5' direction relative to a typical $PVY^{N:O}$, to nt 2307 from nt 2390 position (see Fig 2.7). The answer to the second question may be easier 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).

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, we found only 3 novel recombinant types, ND23 mentioned above that may be termed atypical $PVY^{N:O}$, and two PVY^{C} recombinants, NY110001 and AL100001 with new recombinant structures (Fig. 2.7). 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 (N-Wi) or without (NTNa) non-recombinant isolates suggested that the same types of recombinants, e.g. N-Wi and NTNa, were formed more than once from different parental sequences. Hence, based both on the large number of recombinant types of PVY (Fig. 2.7) and also on the multiple clades characteristic of the same or similar recombinant types, we can 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. The nature of such a limitation or restriction was demonstrated to be unrelated to the physical properties of the PVY RNA around the most common RJs (Hu et al., 2009b), and would be more likely related to some selection pressure provided by the host, perhaps expressing various forms of resistance to the virus.

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), 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.

References

Berger, P.H., Adams, M.J., Barnett, O.W., Brunt, A.A., Hammond, J., Hill, J.H., Jordan, R.L., Kashiwazaki, S., Rybicki, E., Spence, N., Stenger, D.C., Ohki, S.T., Uyeda, I., van Zaayen, A., Valkonen, J., Vetten, H.J., (2005). Family Potyviridae. Virus taxonomy, Classification, Nomenclature of Viruses. Eighth Report of the International Committee on Taxonomy of Viruses. Academic Press, pp. 819–841.

Blanchard, A., Rolland, M., Lacroix, C., Kerlan, C., and Jacquot, E., (2008). Potato virus Y: a century of evolution. Current Topics in Virology, 7: 21-32.

Chikh Ali, M., Gray, S.M., and Karasev, A.V. (2013) An improved multiplex IC-RT-PCR assay distinguishes nine strains of Potato virus Y. Plant Disease 97: 1370-1374.

Chikh-Ali, M., Rowley, J.S., Kuhl, J.C., Gray, S.M., and 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. American Journal of Potato Research, in press.

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.

Cuevas JM, Delaunay A, Visser JC, Bellstedt DU, Jacquot E, et al. (2012) Phylogeography and Molecular Evolution of Potato virus Y. PLoS ONE 7(5): e37853. doi:10.1371/journal.pone.0037853.

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. European J. of Plant Pathology. 126 : 4, 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 Research 32(5), 1792-97.

Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A., (2005). (Eds.), Virus Taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA, USA.

Galvino-Costa, S.B.F., Figueira, A.d.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 Pathology 61(2), 388-398.

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

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16573–582.

Gibbs, A., Ohshima, K. (2010). Potyviruses and the digital revolution. Ann. Rev. Phytopathology, 48: 205-223.

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. Archives of Virology 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 Disease 94, 1384–97.

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.

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. Pot. Res. 87(1), 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. Pot. Res. 90, 7-13.

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.,and 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.

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(21), 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. Archives of Virology 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 16562–563.

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.

Milne, R. G., (1988). Taxonomy of the rod-shaped filamentous viruses. In The Plant Viruses, vol. 4, The Filamentous Plant Viruses, pp. 3-50. Edited by R. G. Milne. New York & London: Plenum Press.

Minin, V., Abdo, Z., Joyce, P., Sullivan, J., (2003). Performance-Based Selection of Likelihood Models for Phylogeny Estimation. Syst. Biol. 52(5), 674-683.

Moury B (2010) A new lineage sheds light on the evolutionary history of Potato virus Y. Molecular Plant Pathology 11: 161–168.

Nikolaeva, O.V., Roop, D., Galvino-Costa, S.F.B., Figueira, A.R., Gray, S.M., and Karasev, A.V. (2012) Epitope mapping for monoclonal antibodies recognizing tuber necrotic strains of Potato virus Y. American Journal of Potato Research 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 Research 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. Journal of Phytopathology 160: 661-673.

Padidam, M., Sawyer, S., Fauquet, C.M., (1999). Possible emergence of new geminiviruses by frequent recombination. Virology 265218–225.

Posada, D., Crandall, K.A., (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. USA 9813757–13762.

Quintero-Ferrer, A., Robles-Hernandez, L., Gonzalez-Franco A.C., Kerlan, C., and Karasev, A.V. (2014) Molecular and biological characterization of a recombinant isolate of Potato virus Y from Mexico. Archives of Virology, published on-line January 9, 2014 (DOI 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.

Romancer, M. L., Kerlan, C. and Nedellec, M., (1994). Biological characterisation of various geographical isolates of potato virus Y inducing superficial necrosis on potato tubers. Plant Pathology 43, 138–144.

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. Retroviruses 11, 1423–1425.

Schubert, J., Fomitcheva, V., Sztangret-Wiśniewska, J., (2007). Differentiation of Potato virus Y strains using improved sets of diagnostic PCR-primers. Journal of Virological Methods 140: 66–74.

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. 34126–129.

Sokal, R., Michener, C., (1958). A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin 38, 1409–1438.

Sukumaran, J., Holder, M.T., (2010). A Python library for phylogenetic computing. Bioinformatics 26, 1569-1571.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28, 2731-2739.

Visser, J.C., Bellstedt, D.U., and Pirie, M.D. (2012). The Recent Recombinant Evolution of a Major Crop Pathogen, Potato virus Y. PLoS ONE 7(11): e50631. doi: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.

Supplementary Information

Table 2.1. Whole genome sequences used for analyses in this work. TBA = to be assigned.

GenBank Isolates			
	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	Unclassified
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	Unclassified
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	C
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-
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

Newly Sequenced Isolates					
	Isolate Name	**GenBank Accession**	**Strain**	**Serology**	**Multiplex**
1	AL100001	TBA	C	O	C
2	CA14	TBA	O	O	O
3	CO11	TBA	O5	O5	O
4	CO86	TBA	O	O	O
5	ID1_4_32B	TBA	O5	O5	O
6	ID_1_7_12B	TBA	O	O	O
7	ID_1258	TBA	O	O	O
8	ID1_1_3A	TBA	N-Wi	O	N-Wi
9	ID1_3_11B	TBA	N-Wi	O	N-Wi
10	ID11_13_11b	TBA	NTNa	N	NTNa
11	ID11_13_12A	TBA	N-Wi	O	N-Wi
12	ID12_102IC3	TBA	NTNa	N	NTNa
13	ID12_110Ban1	TBA	O5	O5	O
14	ID12_22RN8	TBA	NTNa	N	NTNa
15	ID12_401Chf	TBA	N-Wi	O	N-Wi
16	ID125	TBA	N-Wi	O	N-Wi
17	ID1280	TBA	NE-11	N	NE-11
18	ID13_148Oth	TBA	N-Wi	O	N-Wi
19	ID13_610Brw	TBA	NE-11	N	NE-11
20	ID21	TBA	NE-11	N	NE-11
21	ID26	TBA	NE-11	N	NE-11
22	ID281_O5	TBA	O5	O5	O
23	ID38	TBA	NTNa	N	NTNa
24	ID50	TBA	NTNa	N	NTN
25	ID89	TBA	N-Wi	O	N-Wi
26	ID90	TBA	N:O	O	N:O
27	Linda14	TBA	N-Wi	O	Uncl.
28	ME_236_120	TBA	O	O	O
29	ME_236_71	TBA	O	O	O

30	ME10	TBA	NTNa	N	NTNa
31	ME100004	TBA	N:O	O	N:O
32	ME100007	TBA	O	O	O
33	ME100008	TBA	O5	O5	O
34	ME100011	TBA	NTNa	N	NTNa
35	ME100031	TBA	NTNa	N	NTNa
36	ME110008	TBA	NTNa	N	NTNa
37	ME110032	TBA	NTNa	N	NTNa
38	ME200cornell	TBA	O5	O5	O
39	ME4	TBA	NTNa	N	NTNa
40	ME81	TBA	N:O	O	N:O
41	ME9	TBA	NTNa	N	NTNa
42	MI090004	TBA	N:O	O	N:O
43	MI110011	TBA	N-Wi	O	N-Wi
44	MN10c_26	TBA	N:O	O	N:O
45	MN121	TBA	N:O	O	N:O
46	MN13a_39	TBA	N:O	O	N:O
47	MN15_G_52	TBA	N-Wi	O	N-Wi
48	MN21	TBA	N-Wi	O	N-Wi
49	MN85	TBA	NTNa	N	NTNa
50	MSU_45-384a	TBA	Eu-N	N	Eu-N
51	MSU_59-383b	TBA	Eu-N	N	Eu-N
52	MT100006	TBA	Eu-N	N	Eu-N
53	MT100010	TBA	O5	O5	O
54	MT100017	TBA	Eu-N	N	Eu-N
55	MT29	TBA	O5- (no marker)	O	O
56	MT52	TBA	N:O	O	N:O
57	MT63	TBA	O5- (no marker)	O	O
58	ND100040	TBA	NE-11	N	NE-11
59	ND110037	TBA	N:O	O	N:O
60	ND121	TBA	N:O	O	N:O
61	ND18	TBA	N:O	O	N:O
62	ND2	TBA	N-Wi	O	N-Wi
63	ND23	TBA	N:O	O	N:O
64	ND35	TBA	O5	O	O
65	ND65	TBA	N:O	O	N:O
66	ND68	TBA	N:O	O	N:O
67	ND71	TBA	N:O	O	N:O
68	ND98	TBA	N-Wi	O	N-Wi
69	ND99	TBA	N:O	O	N:O
70	NE38	TBA	O	O	O

71	NE40	TBA	N:O	O	N:O
72	NE6	TBA	O	O	O
73	NY090004	TBA	N:O	O	N:O
74	NY090029	TBA	NTNa	N	NTN
75	NY090031	TBA	O	O	O
76	NY100001	TBA	O	O	O
77	NY100002	TBA	O	O	O
78	NY100003	TBA	O	O	O
79	NY100086	TBA	O	O	O
80	NY110001	TBA	Unclassified	O	O
81	NY120001	TBA	N-Wi	O	N-Wi
82	NY120002	TBA	N-Wi	O	N-Wi
83	NY51	TBA	N:O	O	N:O
84	OR16	TBA	N:O	O	N:O
85	OR2	TBA	O	O	O
86	OR20	TBA	O	O	O
87	OR3	TBA	N-Wi	O	N-Wi
88	OR35	TBA	NTNa	N	NTNa
89	Pondo4	TBA	261-4 like	O	Uncl.
90	SU2	TBA	NE-11	N	NE-11
91	WA316	TBA	NE-11	N	NE-11
92	WA70	TBA	N:O	O	N:O
93	WA9	TBA	O	O	O
94	WI120018	TBA	NE-11	N	NE-11
95	WI120092	TBA	O	O	Inconclusive
96	WI120127	TBA	N:O	O	N:O
97	WI3	TBA	O	O	O
98	WI3406	TBA	O	O	O
99	WI62	TBA	N:O	O	N:O
100	WY1	TBA	O5- (no marker)	O	O
101	CO_225	TBA	O5	O5	O
102	CO_28	TBA	O5	O5	O
103	CO12	TBA	O5	O5	O
104	CO238	TBA	O5	O5	O
105	CO254	TBA	O5	O5	O
106	CO275	TBA	O5	O5	O
107	CO32	TBA	O5	O5	O
108	CO39	TBA	O5	O5	O
109	CO53	TBA	O5	O5	O
110	CO55	TBA	O5	O5	O
111	CO6	TBA	O5	O5	O

112	ID_1005	TBA	O5	O5	O
113	ID_236	TBA	O5- (no marker)	O	O
114	ME_222_18	TBA	O5	O5	O
115	ME_250_106	TBA	O5	O5	O
116	ME_250_20	TBA	O5	O5	O
117	ME_323_34	TBA	O	O	O
118	ND127	TBA	O	O	O
119	T1	TBA	O (unusual)	O	O

Table 2.2. Summary of models selected by DT-Modsel 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

Figure Legends

Fig. 2.1. A schematic diagram of previously known and published *Potato virus Y* (PVY) recombinant structures. The ruler at the top represents the PVY genome (ca. 9.7-kb); individual cistrons are presented as rectangles below that with corresponding designations, P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa, NIb, and CP. Potential parental sequences are grouped above the black horizontal dividing bar, with different parents colored differently. Recombinants are below the black bar, and fragments originating from different parents are colored accordingly. The five major recombinant sections, based on common breakpoints, are designated in red numbers and the locations of breakpoints are given below that.

Fig. 2.2. Phylogenetic tree for section 1 (see Fig. 2.1). Different clades and strain types are numbered. Isolates highlighted in red, yellow, and green are unclassified or rare recombinants

(red = previously characterized, green = characterized in this study, yellow = characterized in a separate study).

Fig. 2.3. Phylogenetic tree for section 2 (see Fig. 2.1). Different clades and strain types are numbered. Isolates highlighted in red, yellow, and green are unclassified or rare recombinants (red = previously characterized, green = characterized in this study, yellow = characterized in a separate study). See the Results and Conclusions section for further details.

Fig. 2.4. Phylogenetic tree for section 3 (see Fig. 2.1). Different clades and strain types are numbered. Isolates highlighted in red, yellow, and green are unclassified or rare recombinants (red = previously characterized, green = characterized in this study, yellow = characterized in a separate study). See the Results and Conclusions section for further details.

Fig. 2.5. Phylogenetic tree for section 4 (see Fig. 2.1). Different clades and strain types are numbered. Isolates highlighted in red, yellow, and green are unclassified or rare recombinants (red = previously characterized, green = characterized in this study, yellow = characterized in a separate study). See the Results and Conclusions section for further details.

Fig. 2.6. Phylogenetic tree for section 5 (see Fig. 2.1). Different clades and strain types are numbered. Isolates highlighted in red, yellow, and green are unclassified or rare recombinants (red = previously characterized, green = characterized in this study, yellow = characterized in a separate study).

Fig. 2.7. A schematic diagram like Fig. 2.1. However, this time rare recombinants are added. Those highlighted in red were previously published. Those highlighted in green (Pondo4, NY110001, AL100001, and ND23) were sequenced and their structures determined as part of this study, and represent 3 novel PVY recombinant structures.

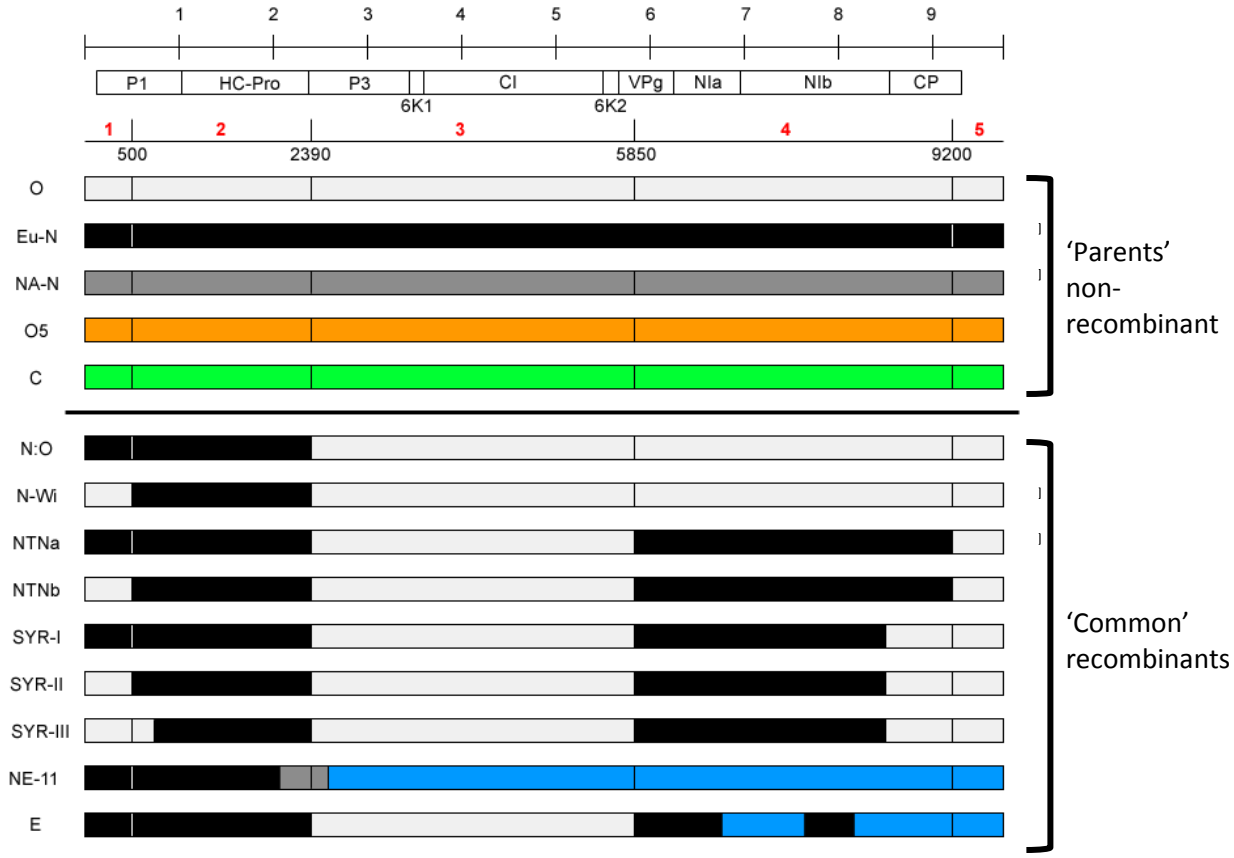
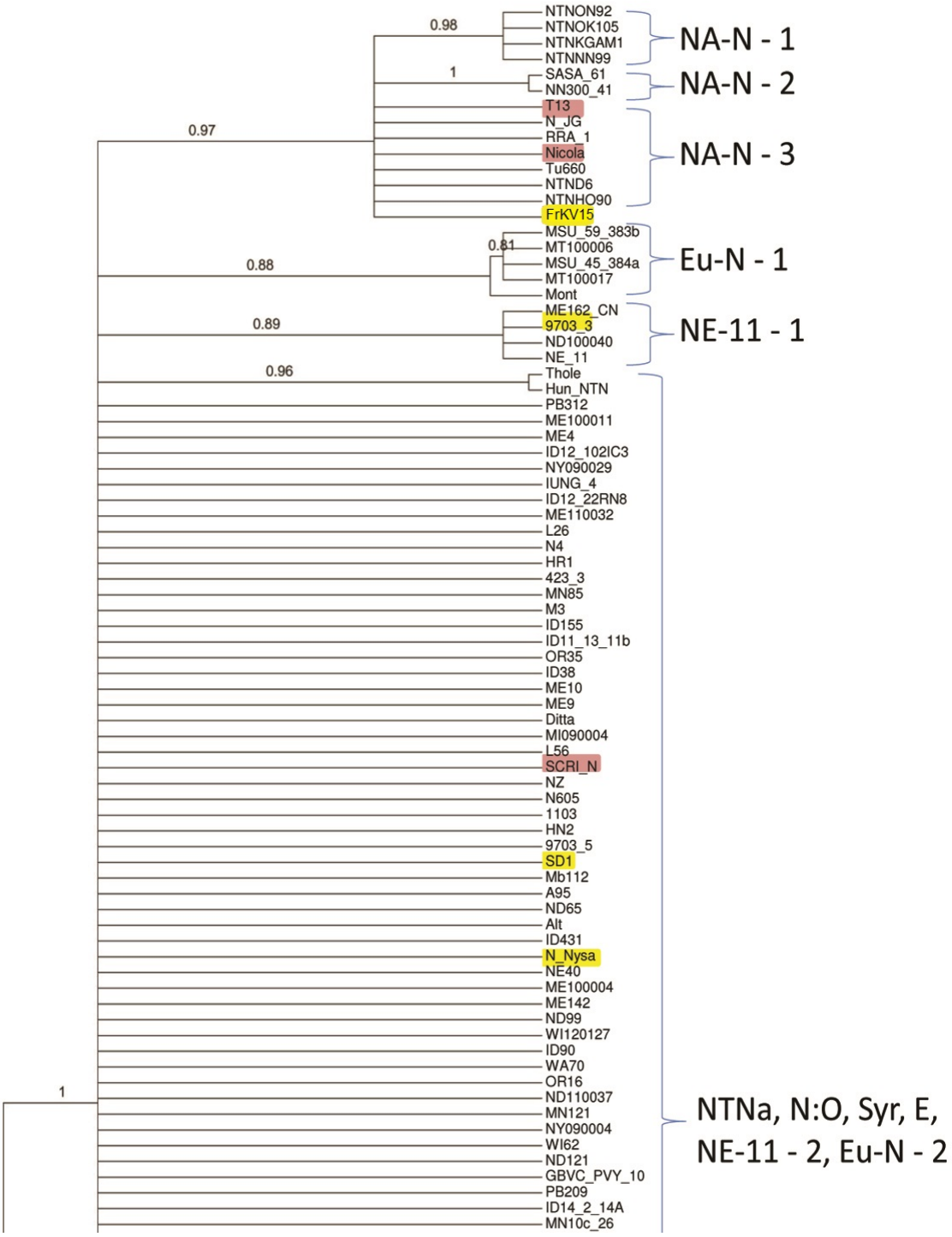
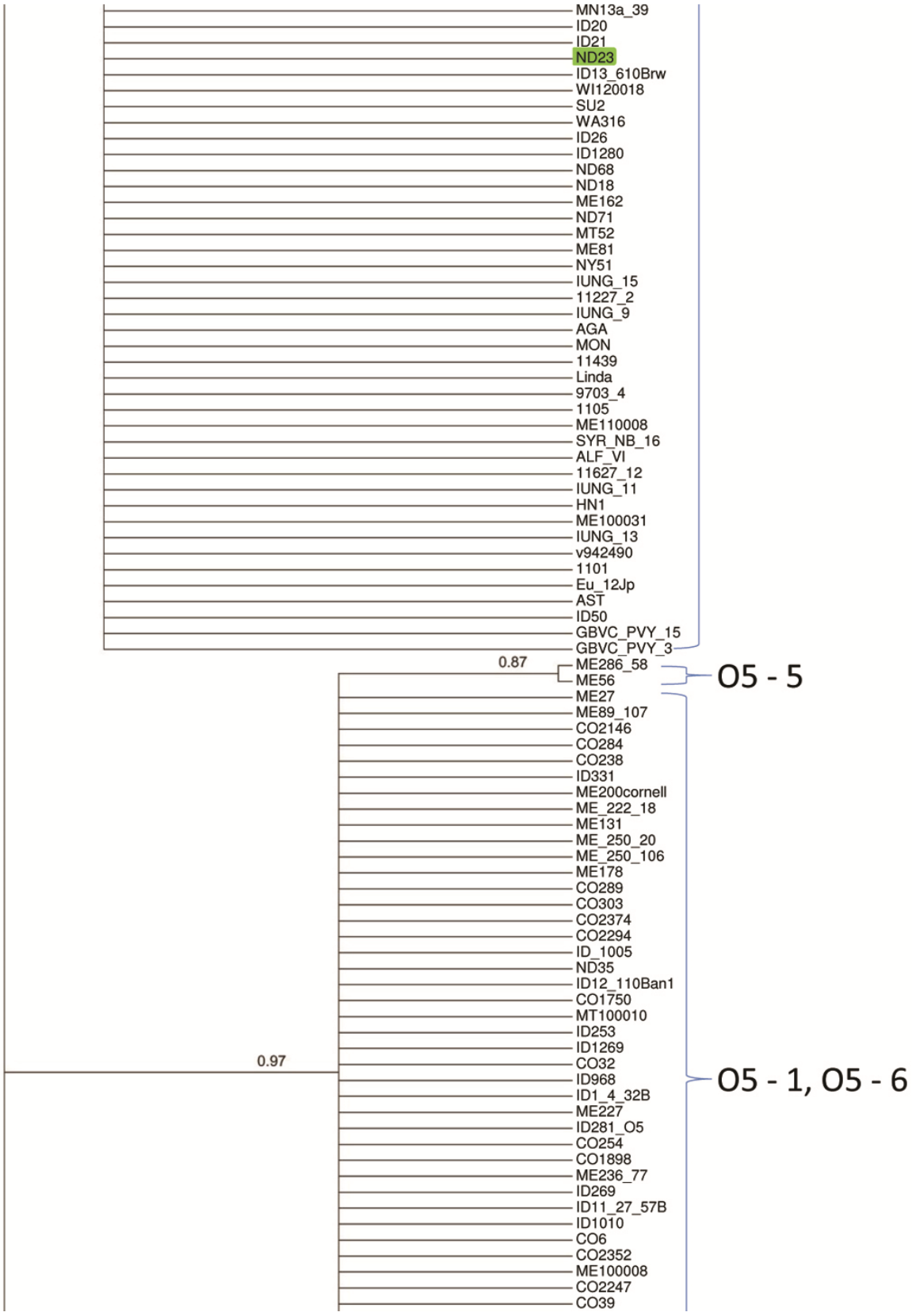
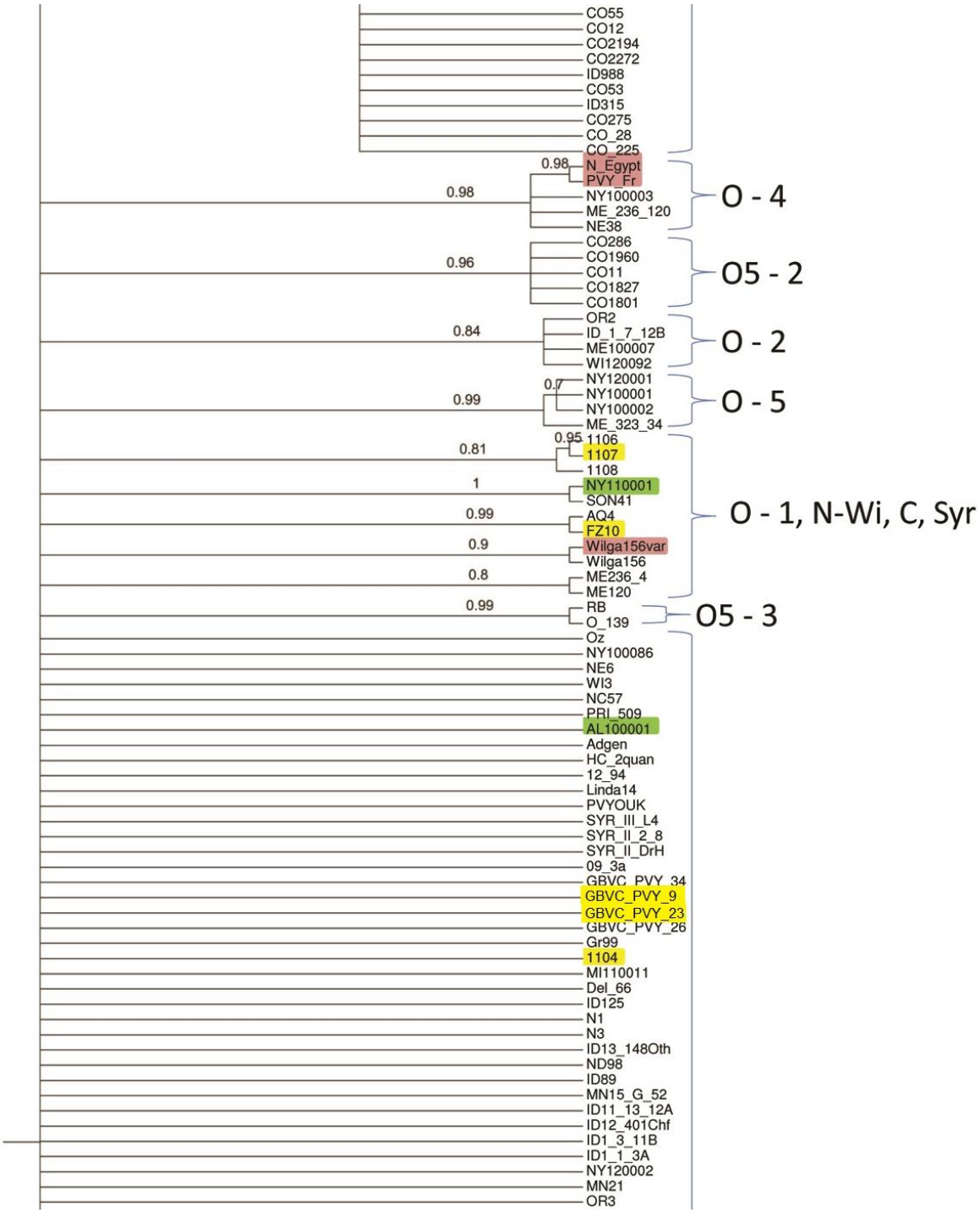

1
2
3
4
5
6
7
8
9
P1
HC-Pro
P3
6K1
CI
6K2
VPg
NIa
NIb
CP
1
2
3
4
5
500
2390
5850
9200
O
Eu-N
NA-N
O5
C
'Parents' non-recombinant
N:O
N-Wi
NTNa
NTNb
SYR-I
SYR-II
SYR-III
NE-11
E
'Common' recombinants

Fig. 2.1


NTNON92
NTNOK105
NTNKGAM1
NTNNN99
SASA_61
NN300_41
T13
N_JG
RRA_1
Nicola
Tu660
NTND6
NTNHO90
FrKV15
MSU_59_383b
MT100006
MSU_45_384a
MT100017
Mont
ME162_CN
9703_3
ND100040
NE_11
Thole
Hun_NTN
PB312
ME100011
ME4
ID12_102IC3
NY090029
IUNG_4
ID12_22RN8
ME110032
L26
N4
HR1
423_3
MN85
M3
ID155
ID11_13_11b
OR35
ID38
ME10
ME9
Ditta
MI090004
L56
SCRI_N
NZ
N605
1103
HN2
9703_5
SD1
Mb112
A95
ND65
Alt
ID431
N_Nysa
NE40
ME100004
ME142
ND99
WI120127
ID90
WA70
OR16
ND110037
MN121
NY090004
WI62
ND121
GBVC_PVY_10
PB209
ID14_2_14A
MN10c_26
0.98
1
0.97
0.81
0.88
0.89
0.96
1
NA-N - 1
NA-N - 2
NA-N - 3
Eu-N - 1
NE-11 - 1
NTNa, N:O, Syr, E,
NE-11 - 2, Eu-N - 2


MN13a_39
ID20
ID21
ND23
ID13_610Brw
WI120018
SU2
WA316
ID26
ID1280
ND68
ND18
ME162
ND71
MT52
ME81
NY51
IUNG_15
11227_2
IUNG_9
AGA
MON
11439
Linda
9703_4
1105
ME110008
SYR_NB_16
ALF_VI
11627_12
IUNG_11
HN1
ME100031
IUNG_13
v942490
1101
Eu_12Jp
AST
ID50
GBVC_PVY_15
GBVC_PVY_3
0.87
ME286_58
ME56
ME27
ME89_107
CO2146
CO284
CO238
ID331
ME200cornell
ME_222_18
ME131
ME_250_20
ME_250_106
ME178
CO289
CO303
CO2374
CO2294
ID_1005
ND35
ID12_110Ban1
CO1750
MT100010
ID253
ID1269
0.97
CO32
ID968
ID1_4_32B
ME227
ID281_O5
CO254
CO1898
ME236_77
ID269
ID11_27_57B
ID1010
CO6
CO2352
ME100008
CO2247
CO39
O5 - 5
O5 - 1, O5 - 6


CO55
CO12
CO2194
CO2272
ID988
CO53
ID315
CO275
CO_28
CO_225
0.98
N_Egypt
PVY_Fr
0.98
NY100003
ME_236_120
NE38
O - 4
CO286
CO1960
0.96
CO11
CO1827
CO1801
O5 - 2
OR2
0.84
ID_1_7_12B
ME100007
WI120092
O - 2
NY120001
0.7
NY100001
0.99
NY100002
ME_323_34
O - 5
0.95
1106
0.81
1107
1108
1
NY110001
SON41
0.99
AQ4
FZ10
0.9
Wilga156var
Wilga156
0.8
ME236_4
ME120
O - 1, N-Wi, C, Syr
0.99
RB
O_139
O5 - 3
Oz
NY100086
NE6
WI3
NC57
PRI_509
AL100001
Adgen
HC_2quan
12_94
Linda14
PVYOUK
SYR_III_L4
SYR_II_2_8
SYR_II_DrH
09_3a
GBVC_PVY_34
GBVC_PVY_9
GBVC_PVY_23
GBVC_PVY_26
Gr99
1104
MI110011
Del_66
ID125
N1
N3
ID13_148Oth
ND98
ID89
MN15_G_52
ID11_13_12A
ID12_401Chf
ID1_3_11B
ID1_1_3A
NY120002
MN21
OR3

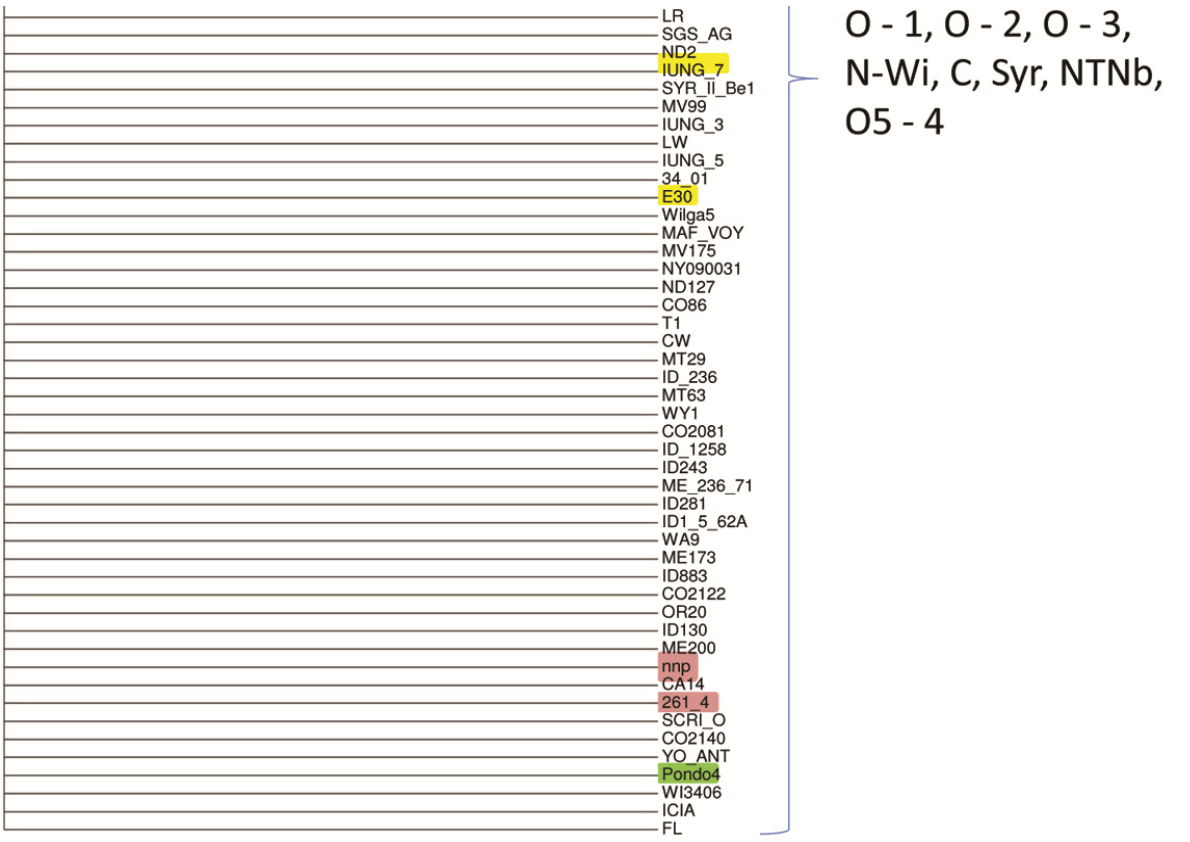
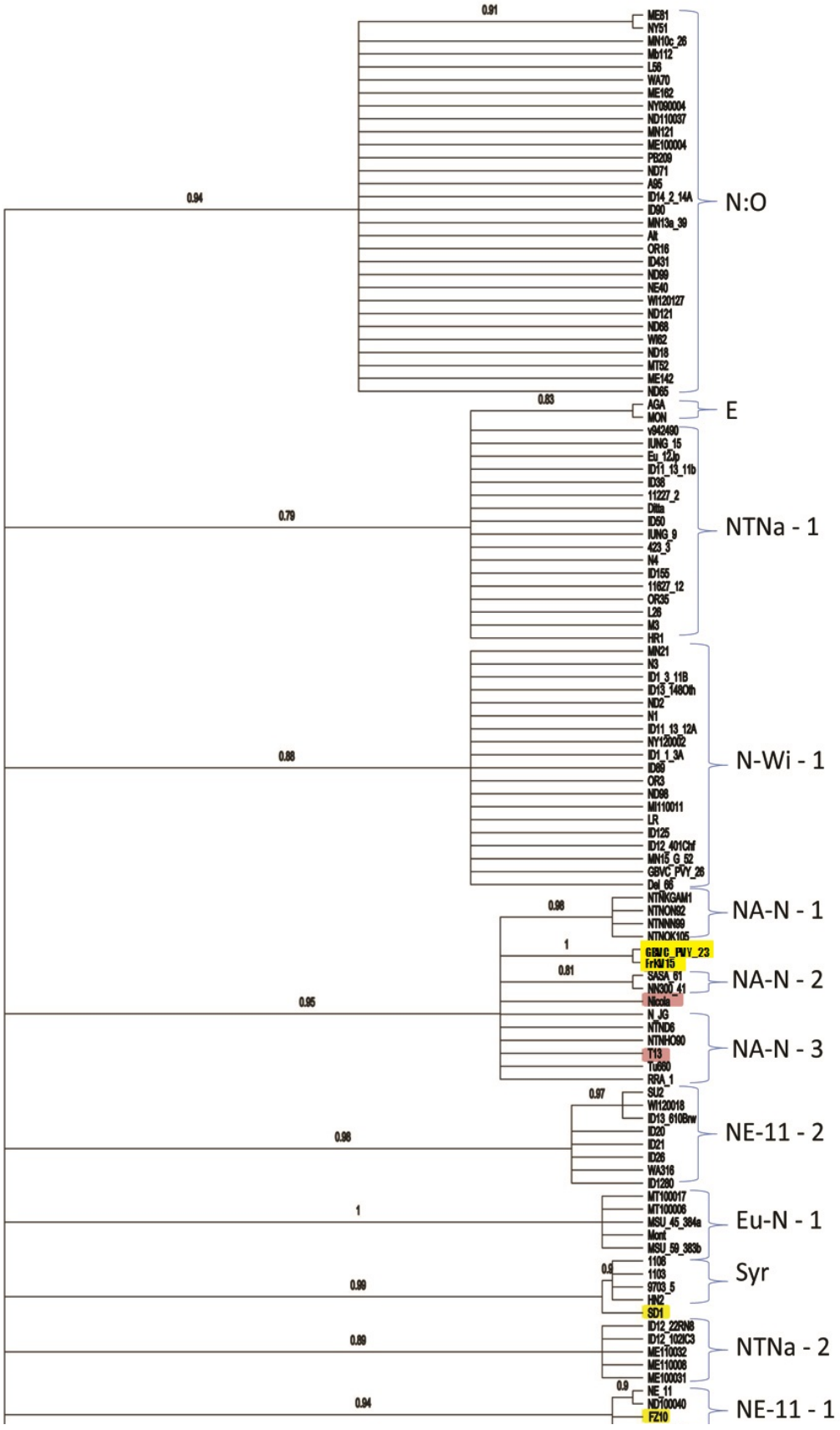
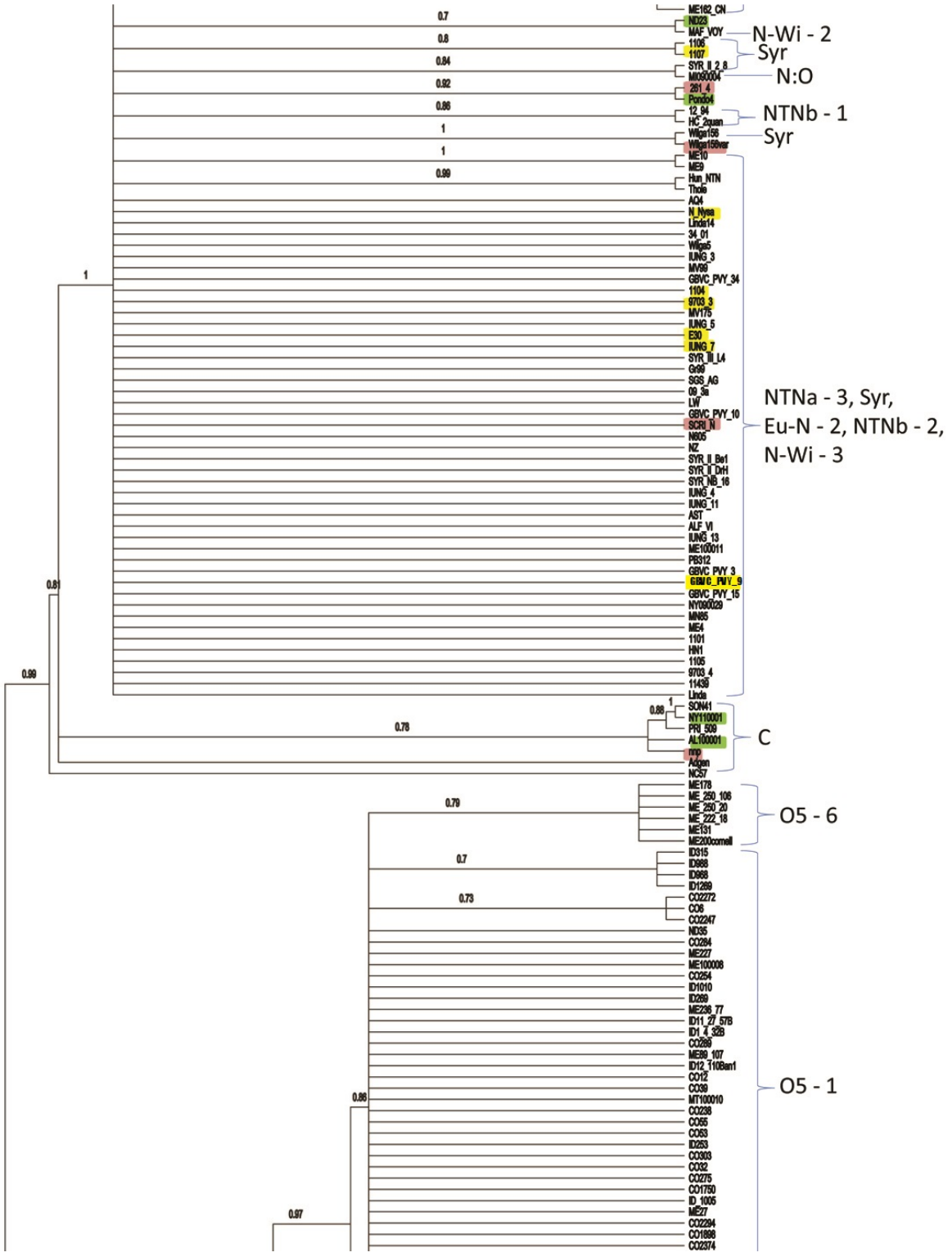

LR
SGS_AG
ND2
IUNG_7
SYR_II_Be1
MV99
IUNG_3
LW
IUNG_5
34_01
E30
Wilga5
MAF_VOY
MV175
NY090031
ND127
CO86
T1
CW
MT29
ID_236
MT63
WY1
CO2081
ID_1258
ID243
ME_236_71
ID281
ID1_5_62A
WA9
ME173
ID883
CO2122
OR20
ID130
ME200
nnp
CA14
261_4
SCRI_O
CO2140
YO_ANT
Pondo4
WI3406
ICIA
FL
O - 1, O - 2, O - 3,
N-Wi, C, Syr, NTNb,
O5 - 4

Fig. 2.2


0.91
0.94
0.83
0.79
0.88
0.98
1
0.81
0.95
0.97
0.98
1
0.99
0.89
0.9
0.94
MEB1
NY51
MN10c_26
Mb112
L56
WA70
ME162
NY080004
ND110037
MN121
ME100004
PB209
ND71
A95
ID14_2_14A
ID90
MN13a_39
Alt
OR16
ID431
ND89
NE40
WI120127
ND121
ND88
WI82
ND18
MT82
ME142
ND65
AGA
MON
v942490
IUNG_15
Eu_12Jp
ID11_13_11b
ID38
11227_2
Ditta
ID50
IUNG_9
423_3
N4
ID155
11627_12
OR35
L26
M3
HR1
MN21
N3
ID1_3_11B
ID13_148Oth
ND2
N1
ID11_13_12A
NY120082
ID1_1_3A
ID69
OR3
ND98
MI110011
LR
ID125
ID12_401Chf
MN15_G_52
GBVC_PVY_26
Del_66
NTNKGAM1
NTNON82
NTNNN69
NTNOK105
GBMC_PVY_23
FrKM15
SASA_61
NN300_41
Nicola
N_JG
NTND6
NTNHO80
T13
Tu660
RRA_1
SU2
WI120018
ID13_610Brw
ID20
ID21
ID26
WA316
ID1280
MT100017
MT100006
MSU_45_384a
Mont
MSU_59_383b
1106
1103
9703_5
HN2
SD1
ID12_22RN8
ID12_102IC3
ME110032
ME110008
ME100031
NE_11
ND100040
FZ10
E
NTNa - 1
N-Wi - 1
NA-N - 1
NA-N - 2
NA-N - 3
NE-11 - 2
Eu-N - 1
Syr
NTNa - 2
NE-11 - 1
N:O


N-Wi - 2
Syr
N:O
NTNb - 1
Syr
NTNa - 3, Syr,
Eu-N - 2, NTNb - 2,
N-Wi - 3
C
O5 - 6
O5 - 1

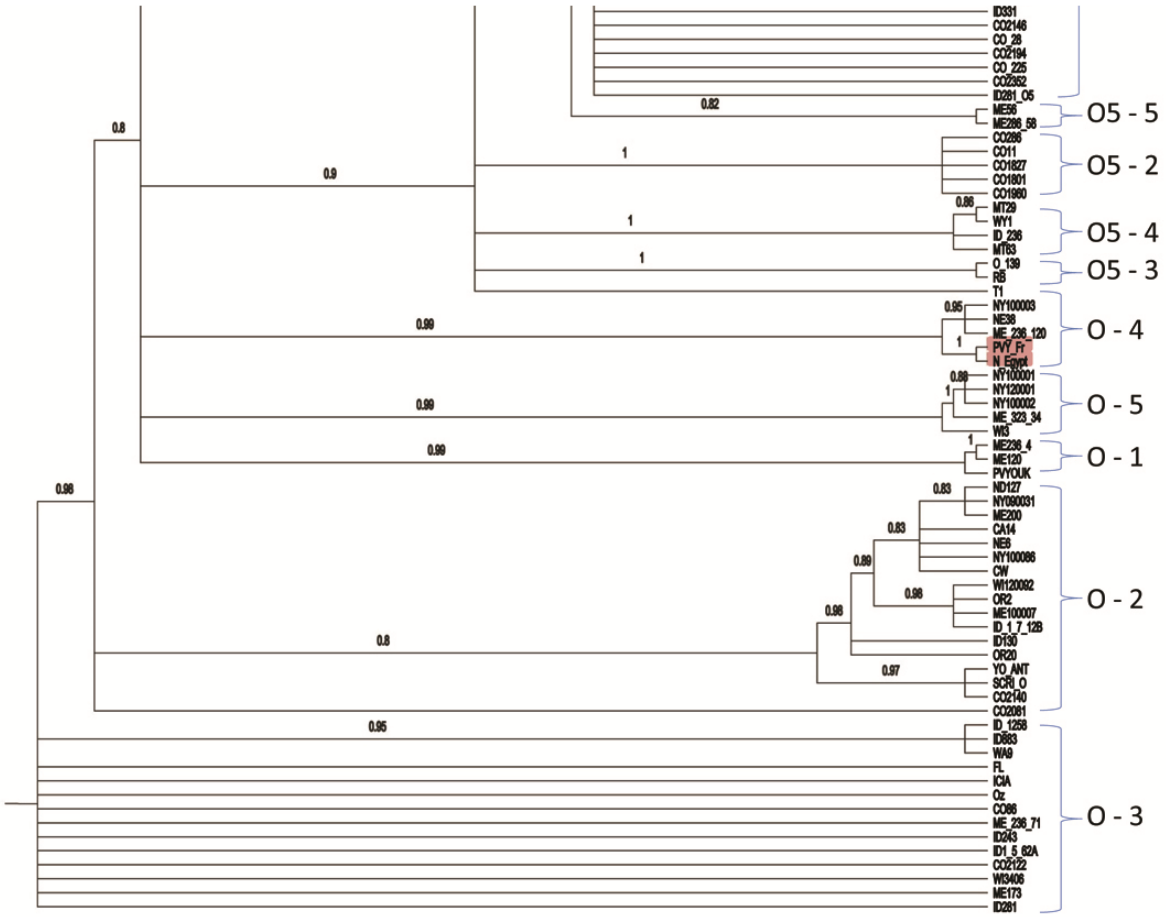
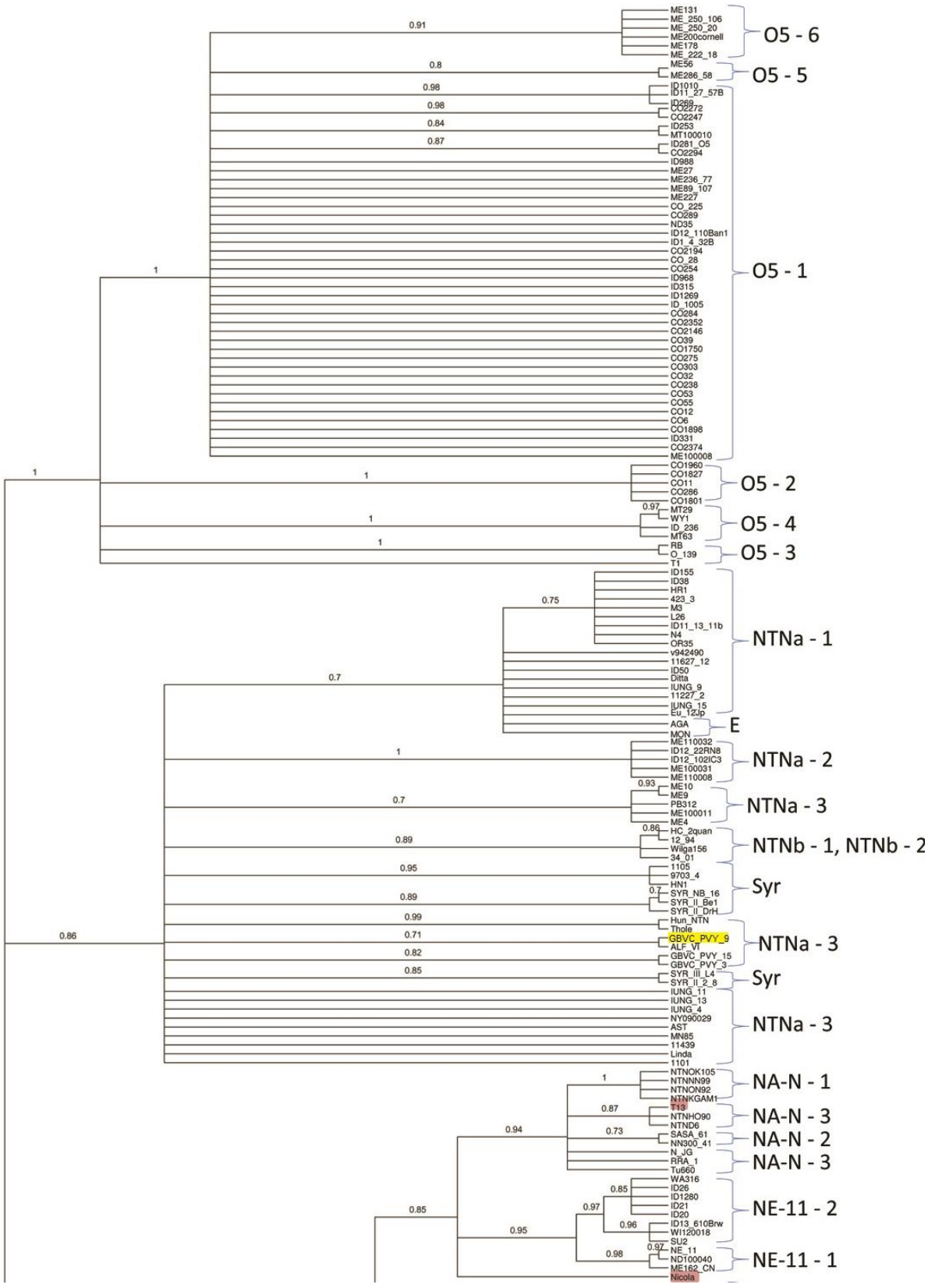

ID331
CO2146
CO_28
CO2194
CO_225
CO2352
ID281_O5
ME56
ME286_56
CO286
CO11
CO1827
CO1801
CO1960
MT29
WY1
ID_236
MT83
O_139
R5
T1
NY100003
NE38
ME_236_120
PVY_Fr
N_Egypt
NY100001
NY120001
NY100002
ME_323_34
WI3
ME236_4
ME120
PVYOUK
ND127
NY090031
ME200
CA14
NE6
NY100086
CW
WI120092
OR2
ME100007
ID_1_7_12B
ID130
OR20
YO_ANT
SCRI_O
CO2140
CO2081
ID_1258
ID683
WA9
FL
IC1A
Oz
CO86
ME_236_71
ID243
ID1_5_62A
CO2122
WI3406
ME173
ID281
O5 - 5
O5 - 2
O5 - 4
O5 - 3
O - 4
O - 5
O - 1
O - 2
O - 3

Fig. 2.3


ME131
ME_250_106
ME_250_20
ME200cornell
ME178
ME_222_18
ME56
ME286_58
ID1010
ID11_27_57B
ID269
CO2272
CO2247
ID253
MT100010
ID281_O5
CO2294
ID988
ME27
ME236_77
ME89_107
ME227
CO_225
CO289
ND35
ID12_110Ban1
ID1_4_32B
CO2194
CO_28
CO254
ID968
ID315
ID1269
ID_1005
CO284
CO2352
CO2146
CO39
CO1750
CO275
CO303
CO32
CO238
CO53
CO55
CO12
CO6
CO1898
ID331
CO2374
ME100008
CO1960
CO1827
CO11
CO286
CO1801
MT29
WY1
ID_236
MT63
RB
O_139
T1
ID155
ID38
HR1
423_3
M3
L26
ID11_13_11b
N4
OR35
v942490
11627_12
ID50
Ditta
IUNG_9
11227_2
IUNG_15
Eu_12Jp
AGA
MON
ME110032
ID12_22RN8
ID12_102IC3
ME100031
ME110008
ME10
ME9
PB312
ME100011
ME4
HC_2quan
12_94
Wilga156
34_01
1105
9703_4
HN1
SYR_NB_16
SYR_II_Be1
SYR_II_DrH
Hun_NTN
Thole
GBVC_PVY_9
ALF_VT
GBVC_PVY_15
GBVC_PVY_3
SYR_III_L4
SYR_II_2_8
IUNG_11
IUNG_13
IUNG_4
NY090029
AST
MN85
11439
Linda
1101
NTNOK105
NTNNN99
NTNON92
NTNKGAM1
T13
NTNHO90
NTND6
SASA_61
NN300_41
N_JG
RRA_1
Tu660
WA316
ID26
ID1280
ID21
ID20
ID13_610Brw
WI120018
SU2
NE_11
ND100040
ME162_CN
Nicola
O5 - 6
O5 - 5
O5 - 1
O5 - 2
O5 - 4
O5 - 3
NTNa - 1
E
NTNa - 2
NTNa - 3
NTNb - 1, NTNb - 2
Syr
NTNa - 3
Syr
NTNa - 3
NA-N - 1
NA-N - 3
NA-N - 2
NA-N - 3
NE-11 - 2
NE-11 - 1

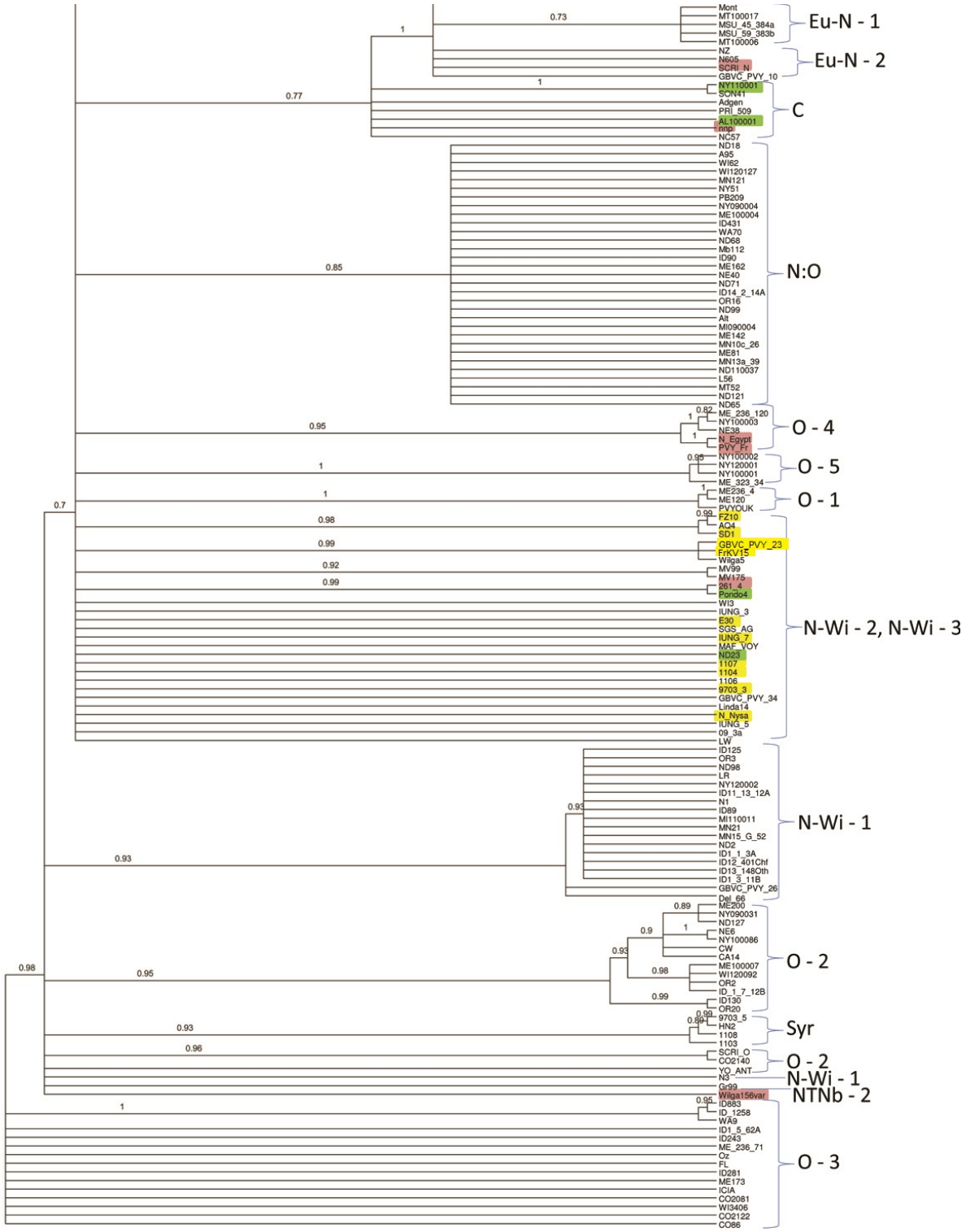
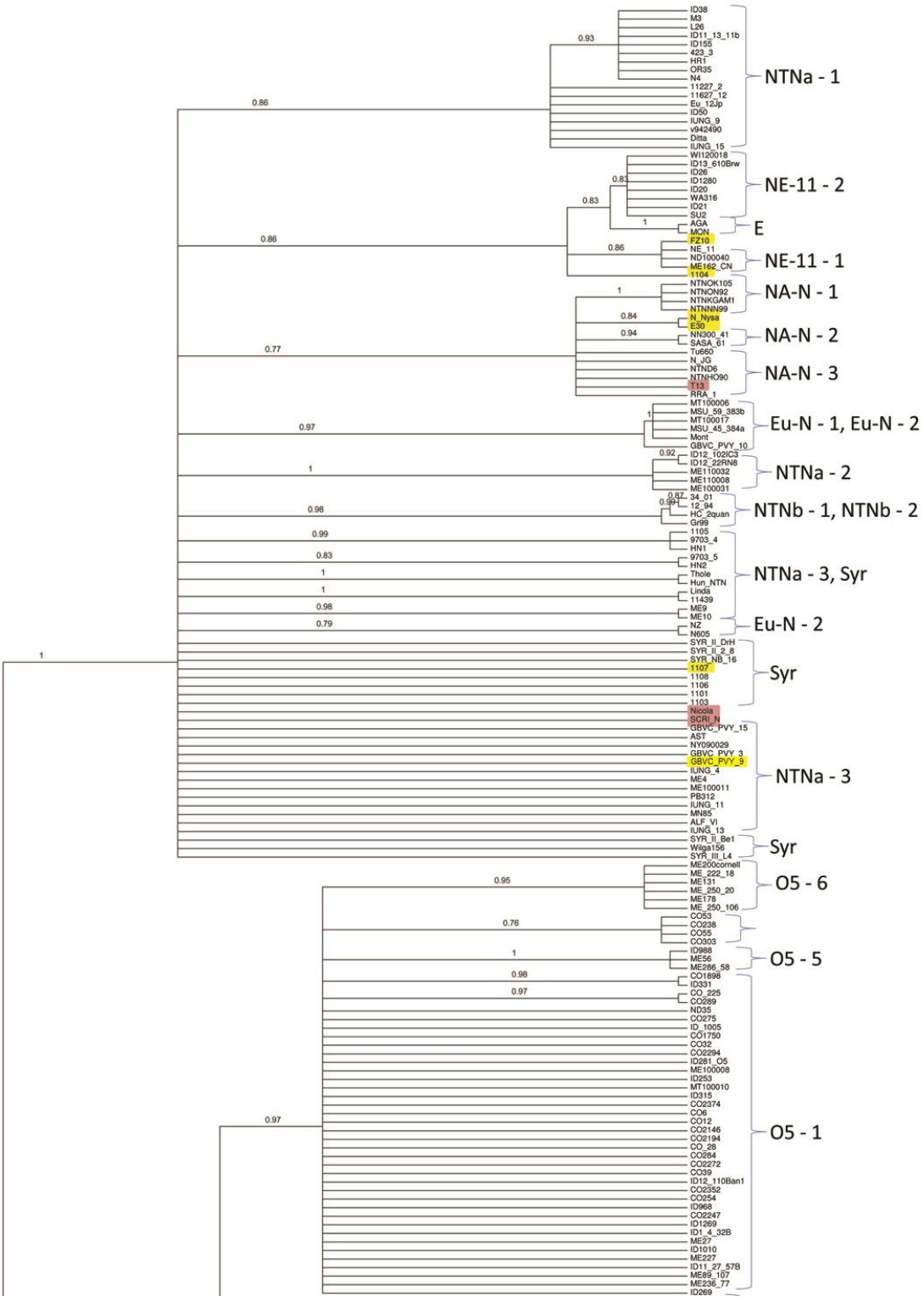


Fig. 2.4


NTNa - 1
NE-11 - 2
E
NE-11 - 1
NA-N - 1
NA-N - 2
NA-N - 3
Eu-N - 1, Eu-N - 2
NTNa - 2
NTNb - 1, NTNb - 2
NTNa - 3, Syr
Eu-N - 2
Syr
NTNa - 3
Syr
O5 - 6
O5 - 5
O5 - 1

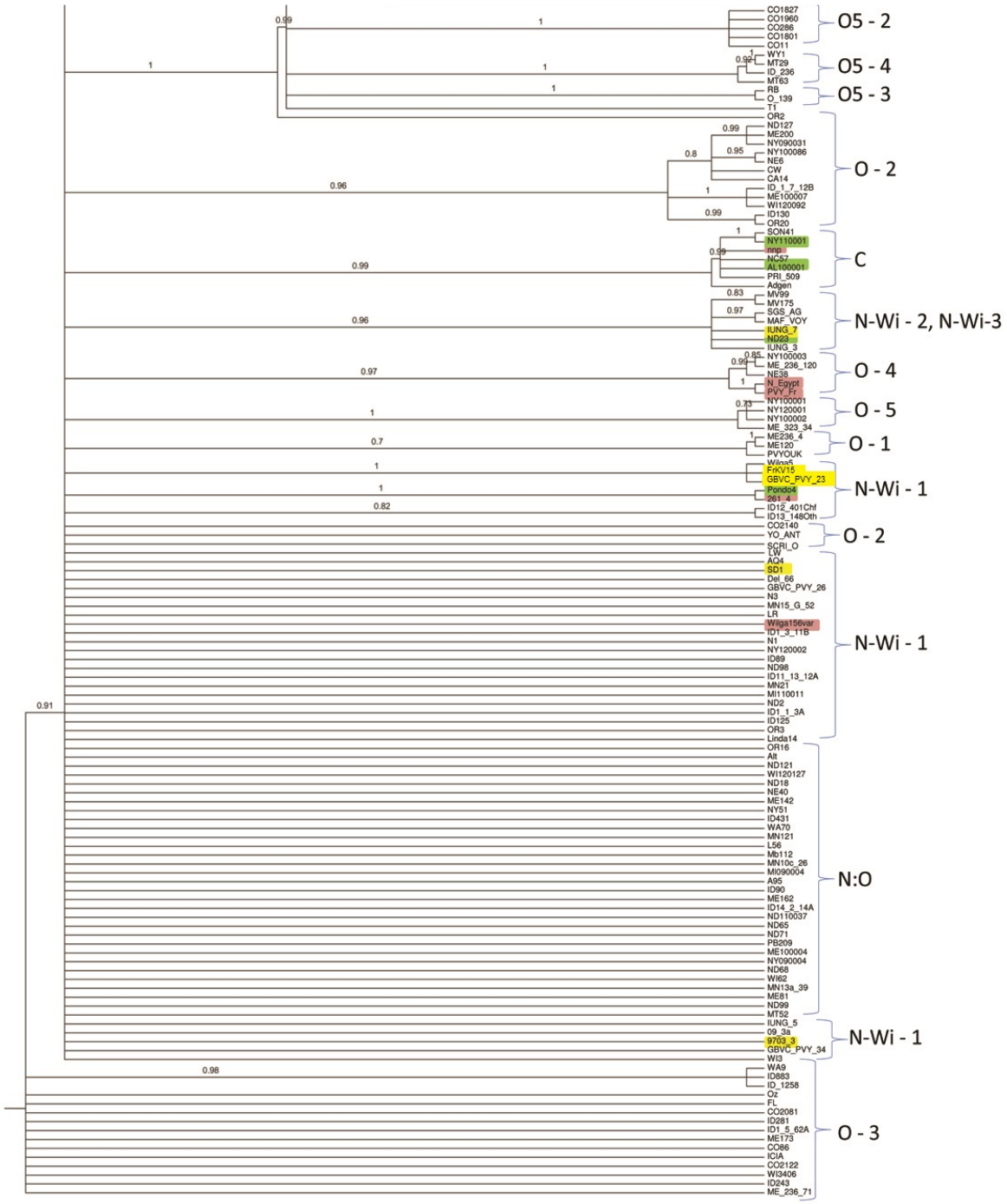
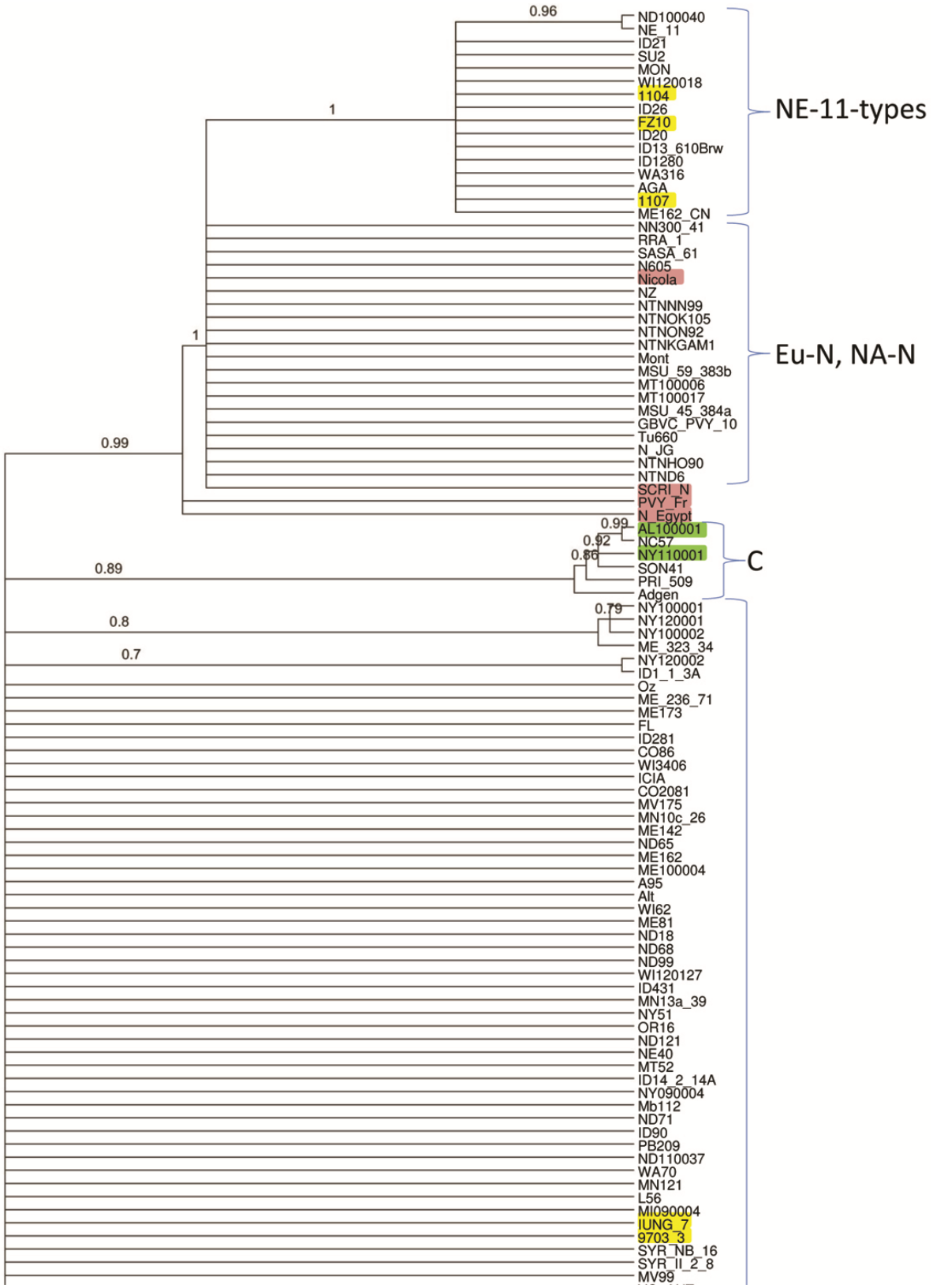
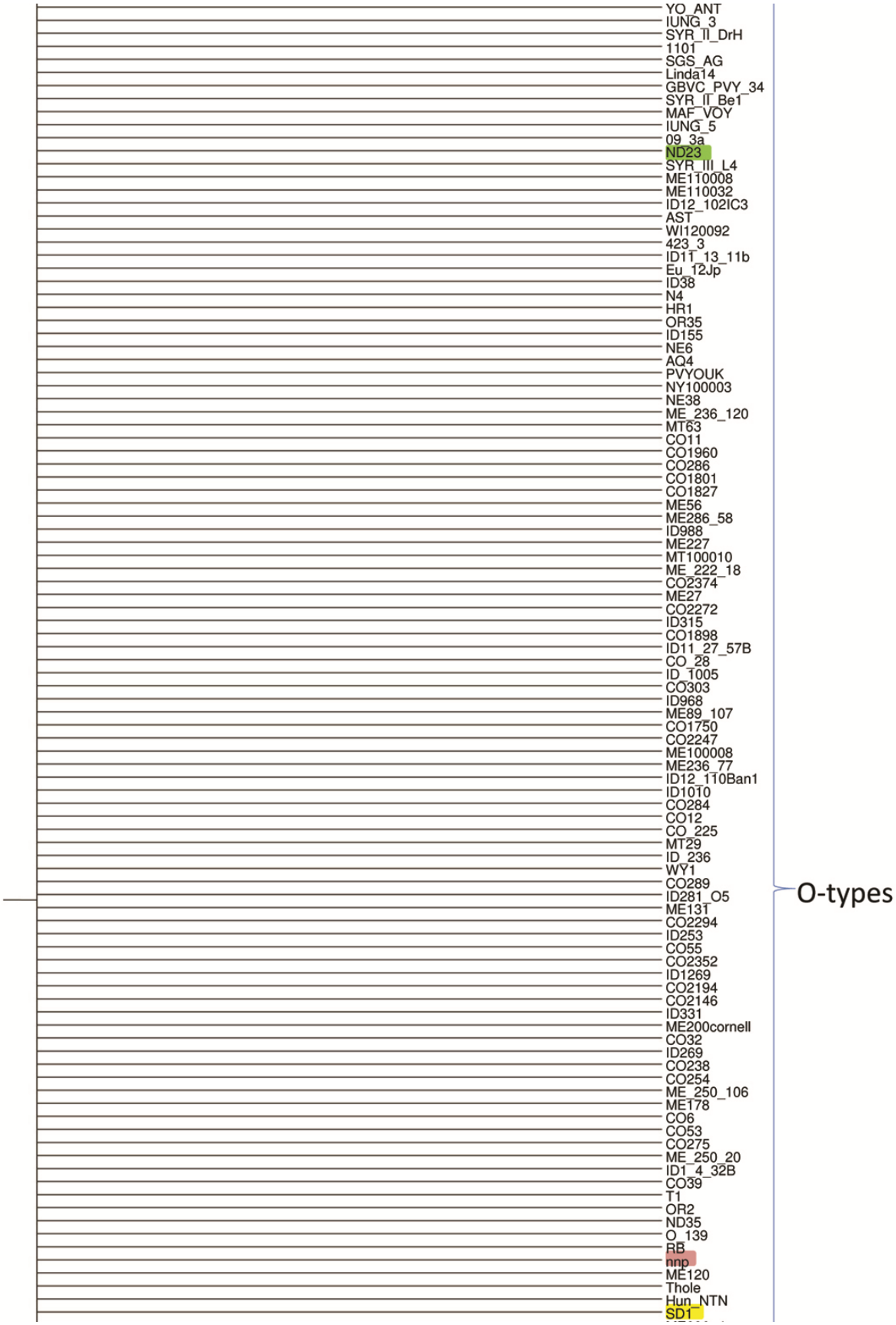

CO1827
CO1960
CO286
CO1801
CO11
WY1
MT29
ID_236
MT63
RB
O_139
T1
OR2
ND127
ME200
NY090031
NY100086
NE6
CW
CA14
ID_1_7_12B
ME100007
WI120092
ID130
OR20
SON41
NY110001
mnp_i
NC57
AL100001
PRI_509
Adgen
MV99
MV175
SGS_AG
MAF_VOY
IUNG_7
ND23
IUNG_3
NY100003
ME_236_120
NE38
N_Egypt
PVY_Fr
NY100001
NY120001
NY100002
ME_323_34
ME236_4
ME120
PVYOUK
Wilga5
FrKV15
GBVC_PVY_23
Pondo4
261_4
ID12_401Chf
ID13_148Oth
CO2140
YO_ANT
SCRI_O
LW
AQ4
SD1
Del_66
GBVC_PVY_26
N3
MN15_G_52
LR
Wilga156var
ID1_3_11B
N1
NY120002
ID89
ND98
ID11_13_12A
MN21
MI110011
ND2
ID1_1_3A
ID125
OR3
Linda14
OR16
Alt
ND121
WI120127
ND18
NE40
ME142
NY91
ID431
WA70
MN121
L56
Mb112
MN10c_26
MI090004
A95
ID90
ME162
ID14_2_14A
ND110037
ND65
ND71
PB209
ME100004
NY090004
ND68
WI82
MN13a_39
ME81
ND99
MT52
IUNG_5
09_3a
9709_3
GBVC_PVY_34
WI3
WA9
ID883
ID_1258
Oz
FL
CO2081
ID281
ID1_5_62A
ME173
CO86
ICIA
CO2122
WI3406
ID243
ME_236_71
O5 - 2
O5 - 4
O5 - 3
O - 2
C
N-Wi - 2, N-Wi-3
O - 4
O - 5
O - 1
N-Wi - 1
O - 2
N-Wi - 1
N:O
N-Wi - 1
O - 3

Fig. 2.5


0.96
1
1
0.99
0.99
0.92
0.86
0.89
0.79
0.8
0.7
ND100040
NE_11
ID21
SU2
MON
WI120018
1104
ID26
FZ10
ID20
ID13_610Brw
ID1280
WA316
AGA
1107
ME162_CN
NN300_41
RRA_1
SASA_61
N605
Nicola
NZ
NTNNN99
NTNOK105
NTNON92
NTNKGAM1
Mont
MSU_59_383b
MT100006
MT100017
MSU_45_384a
GBVC_PVY_10
Tu660
N_JG
NTNHO90
NTND6
SCRI_N
PVY_Fr
N_Egypt
AL100001
NC57
NY110001
SON41
PRI_509
Adgen
NY100001
NY120001
NY100002
ME_323_34
NY120002
ID1_1_3A
Oz
ME_236_71
ME173
FL
ID281
CO86
WI3406
ICIA
CO2081
MV175
MN10c_26
ME142
ND65
ME162
ME100004
A95
Alt
WI62
ME81
ND18
ND68
ND99
WI120127
ID431
MN13a_39
NY51
OR16
ND121
NE40
MT52
ID14_2_14A
NY090004
Mb112
ND71
ID90
PB209
ND110037
WA70
MN121
L56
MI090004
IUNG_7
9703_3
SYR_NB_16
SYR_II_2_8
MV99
NE-11-types
Eu-N, NA-N
C


YO_ANT
IUNG_3
SYR_II_DrH
1101
SGS_AG
Linda14
GBVC_PVY_34
SYR_II_Be1
MAF_VOY
IUNG_5
09_3a
ND23
SYR_III_L4
ME110008
ME110032
ID12_102IC3
AST
WI120092
423_3
ID11_13_11b
Eu_12Jp
ID38
N4
HR1
OR35
ID155
NE6
AQ4
PVYOUK
NY100003
NE38
ME_236_120
MT63
CO11
CO1960
CO286
CO1801
CO1827
ME56
ME286_58
ID988
ME227
MT100010
ME_222_18
CO2374
ME27
CO2272
ID315
CO1898
ID11_27_57B
CO_28
ID_1005
CO303
ID968
ME89_107
CO1750
CO2247
ME100008
ME236_77
ID12_110Ban1
ID1010
CO284
CO12
CO_225
MT29
ID_236
WY1
CO289
ID281_O5
ME131
CO2294
ID253
CO55
CO2352
ID1269
CO2194
CO2146
ID331
ME200cornell
CO32
ID269
CO238
CO254
ME_250_106
ME178
CO6
CO53
CO275
ME_250_20
ID1_4_32B
CO39
T1
OR2
ND35
O_139
RB
nnp
ME120
Thole
Hun_NTN
SD1
O-types

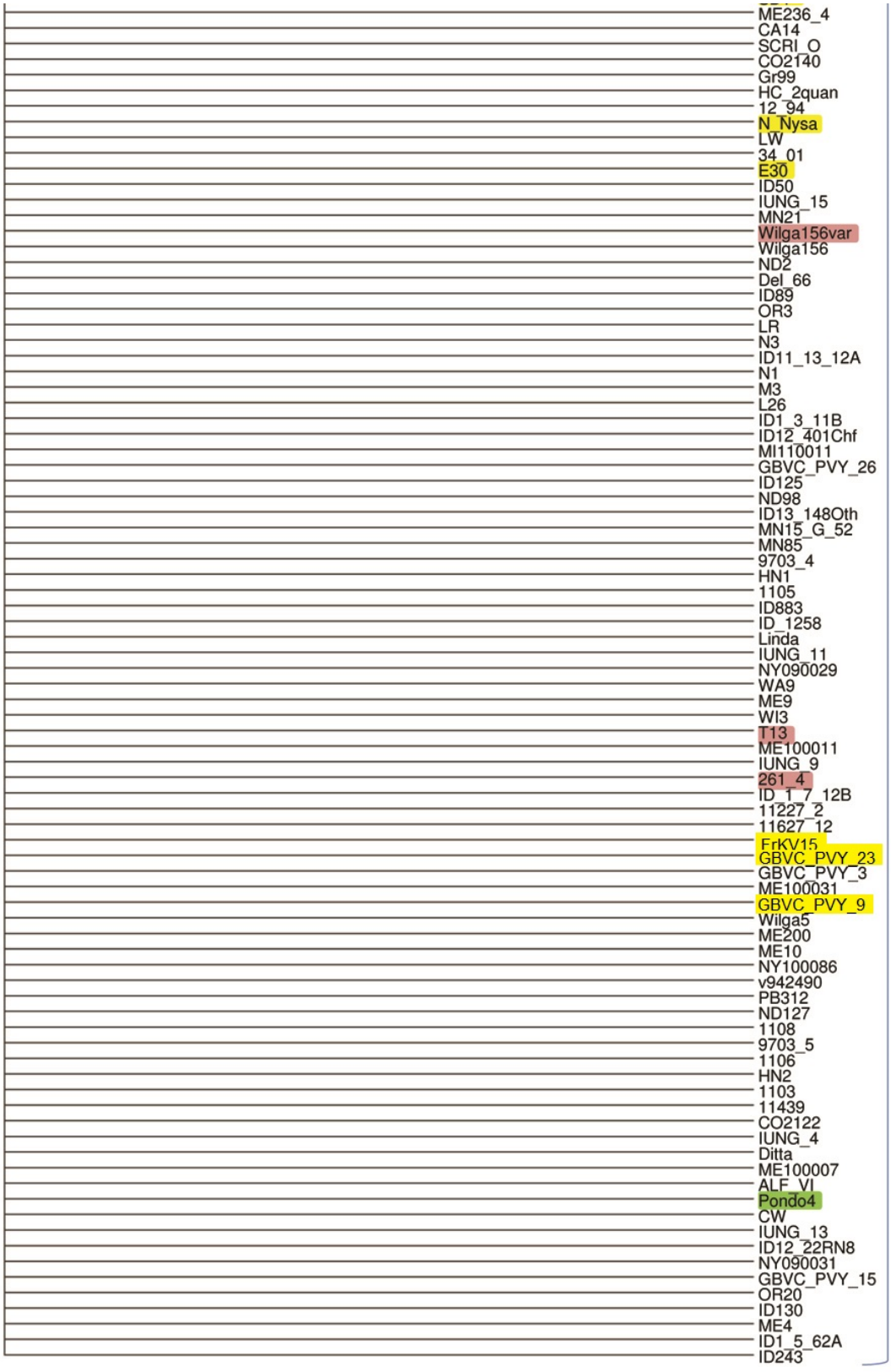

ME236_4
CA14
SCRI_O
CO2140
Gr99
HC_2quan
12_94
N_Nysa
LW
34_01
E30
ID50
IUNG_15
MN21
Wilga156var
Wilga156
ND2
Del_66
ID89
OR3
LR
N3
ID11_13_12A
N1
M3
L26
ID1_3_11B
ID12_401Chf
MI110011
GBVC_PVY_26
ID125
ND98
ID13_148Oth
MN15_G_52
MN85
9703_4
HN1
1105
ID883
ID_1258
Linda
IUNG_11
NY090029
WA9
ME9
WI3
T13
ME100011
IUNG_9
261_4
ID_1_7_12B
11227_2
11627_12
FrKV15
GBVC_PVY_23
GBVC_PVY_3
ME100031
GBVC_PVY_9
Wilga5
ME200
ME10
NY100086
v942490
PB312
ND127
1108
9703_5
1106
HN2
1103
11439
CO2122
IUNG_4
Ditta
ME100007
ALF_VI
Pondo4
CW
IUNG_13
ID12_22RN8
NY090031
GBVC_PVY_15
OR20
ID130
ME4
ID1_5_62A
ID243

Fig. 2.6

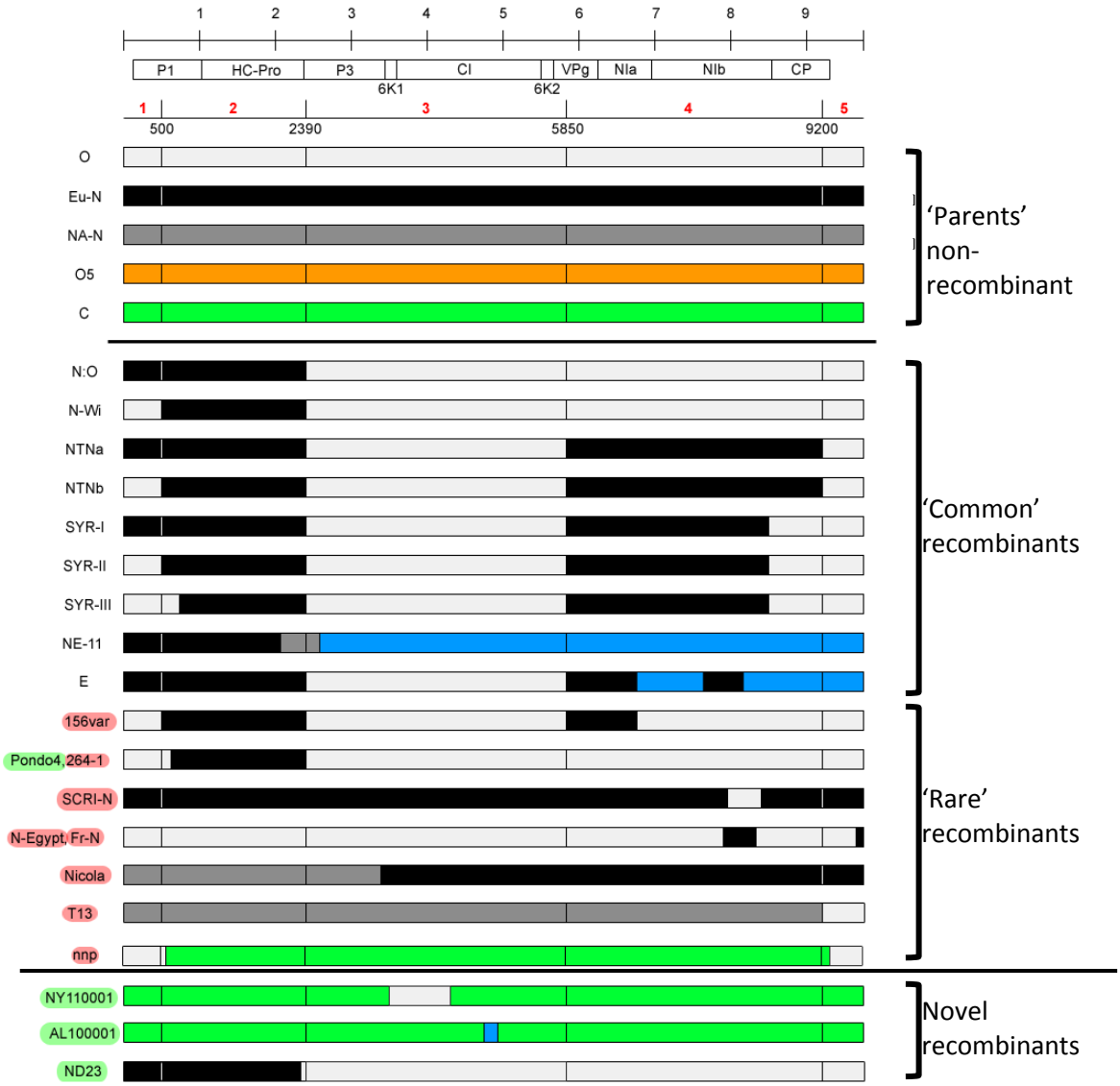

1
2
3
4
5
6
7
8
9
P1
HC-Pro
P3
6K1
CI
6K2
VPg
NIa
NIb
CP
1
2
3
4
5
500
2390
5850
9200
O
Eu-N
NA-N
O5
C
'Parents' non-recombinant
N:O
N-Wi
NTNa
NTNb
SYR-I
SYR-II
SYR-III
NE-11
E
'Common' recombinants
156var
Pondo4,264-1
SCRI-N
N-Egypt,Fr-N
Nicola
T13
nnp
'Rare' recombinants
NY110001
AL100001
ND23
Novel recombinants

Fig. 2.7

CHAPTER 3: Phylogenetic and Recombinant Analysis Reveals Nine novel Recombinants of *Potato virus Y*

Kelsie J Evans[1,2] and Alexander V. Karasev[1,2]

[1]*Department of PSES and* [2]*Bioinformatics and Computational Biology Program, University of Idaho, Moscow, ID*

Abstract

Potato virus Y (PVY) exists as a complex of strains and genetic variants including sixteen currently described recombinants. The majority of these PVY recombinants are composed of genome segments coming from PVY^{O} and PVY^{N} parental genomes, and only five recombinant types include segments from other parental genomes, e.g. PVY^{C}, $PVY^{NA\text{-}N}$, and $PVY^{NE\text{-}11}$. In the course of phylogenetic analyses of a large collection of 285 whole genomes of PVY isolates, ten isolates were found to be inconsistent with their initial strain typing. When subjected to a thorough recombination analysis using programs from the RDP4 package, PVY isolates FrKV15, GBVC_ 23, GBVC_ 9, 9703-3, SD1, 1104, FZ10, 1107, E30, and N-Nysa were found to represent nine novel PVY recombinant structures not reported before. Two novel recombinants, FrKV15 and GBVC_ 23, resembled $PVY^{N:O}$ and $PVY^{N\text{-}Wi}$ strains, respectively, but contained segments of $PVY^{NA\text{-}N}$ sequences instead of PVY^{N}. Six of the nine novel structures were composed of segments originating from three different parental sequences, PVY^{O}, PVY^{N}, and $PVY^{NA\text{-}N}$ or $PVY^{NE\text{-}11}$. PVY isolate 1107 had a PVY^{O} segment inserted in the capsid protein cistron, changing the serotype of the isolate. All these new recombinant structures may present a challenge for correct PVY isolate detection and typing to strain.

Introduction

Potato virus Y (PVY) is a serious problem in potato production worldwide, substantially affecting both tuber yield and quality (Gray et al. 2010; Karasev and Gray, 2013). PVY is a type member of the genus *Potyvirus*, family *Potyviridae*, and has a positive, single-stranded RNA genome coding for a single polyprotein that is later processed into mature virus proteins by virus-specific proteases (Berger et al., 2005). PVY exists as a complex of strains that may be defined genetically, based on hypersensitive resistance (HR) reactions towards three known *N* genes in potato (Jones, 1990; Singh et al., 2008; Chikh-Ali et al., 2014), or molecularly, based on genome sequences and recombination patterns (Karasev and Gray, 2013; Chikh-Ali et al., 2013). Currently, nine strains of PVY have been defined (Karasev and Gray, 2013), with at least sixteen recombinants and genome variants reported (Schubert et al., 2007; Lorenzen et al., 2008; Galvino-Costa et al., 2012a,b; Ogawa et al., 2012; Karasev and Gray, 2013; Chikh-Ali et al., 2013).

Potyviruses are well-known for their propensity to rapid evolution through accumulation of mutations and even more rapidly through recombination, where different strains of the virus exchange large segments of their genome in order to gain selective advantages in a particular host or in different environments (Gibbs and Ohshima, 2010). In an earlier study, the number of recombinant patterns characteristic of various PVY recombinants around the globe was found relatively limited, with parents represented mostly by PVY^{O} and PVY^{N} sequences (Hu et al., 2009). Nevertheless, additional types of PVY recombinants were found with genome sequences other than O or N sequences – a new recombinant NE-11 was found carrying sequences of strains N, NA-N, and NE-11 (Lorenzen et al., 2008), and recombinant PVY^{E} carrying sequences of strains N, O, and NE-11 (Galvino-Costa et al.,

2012a). Biological significance of these recombination events between PVY strains is not yet fully understood, although some recombinants, like PVY^{NTN} and PVY^{E}, were characterized on a set of potato indicators carrying different resistance genes (Kerlan et al., 2011; Galvino-Costa et al., 2012b; Quintero-Ferrer et al., 2014). It was hypothesized that PVY recombination may be driven by the need of the virus to overcome various resistance genes conferring strain-specific HR in potato (Kerlan et al., 2011; Karasev and Gray, 2013). In addition, a substantial genetic diversity was revealed among PVY isolates belonging to the same strain currently defined as PVY^{O} (Karasev et al., 2010, 2011; Ogawa et al., 2012) or PVY^{NA-N} (Ogawa et al., 2012) that allowed us to address the origin and evolution of PVY recombinants carrying segments of these parental sequences.

Recently, a very large number of whole genomes of PVY from different strains was sequenced and subjected to a similar analysis of genome diversity in order to determine origins of several major recombinants of PVY found in potato through examination of phylogenies for O- and N-specific segments common in both recombinant and non-recombinant PVY genomes (Evans et al., 2014). However, besides establishing phylogenetic relationships between common genome segments, and revealing possible parental lineages from non-recombinant PVY genomes (Evans et al., 2014), this analysis also identified ten unusual PVY sequences, all from the GenBank database, with potentially novel recombination structures. Here, we present a thorough investigation of these ten PVY sequences, for isolates FrKV15, GBVC_ 23, GBVC_ 9, 9703-3, SD1, 1104, FZ10, 1107, E30, and N-Nysa, and the resulting assignment of these to nine novel recombination patterns not reported previously for any PVY isolate. Existence of these novel recombinants creates

additional challenges for PVY strain detection and differentiation, and requires additional tools to correctly type PVY isolates to strain.

Materials and Methods

Sequence Dataset and Phylogenetic Analysis

A total of 285 whole-genome PVY sequences was used for analyses (Table 3.1). The dataset included 166 sequences extracted from the GenBank database and 119 recently sequenced PVY genomes (Evans et al., 2014).

Sequence alignment 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 the isolates to strain, and strain types were checked against RT-PCR and serological data, when available, and additionally checked with BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Sokal and Michener, 1958; Martin et al., 2010). Genomes were then all divided into the 5 major recombinant sections as described (Evans et al., 2014) using Seqret, and model selection was run on each section using DT-ModSel (Rice et al., 2000; Minin et al., 2003). A hundred maximum likelihood (ML) trees were constructed in Garli using the model information acquired in DTModSel, with each tree run for 3,000,000 generations or until there were no longer improvements between generations (Zwickl, 2006). The very best tree of all the runs was selected as the reference tree and Sumtrees was used to create bootstraps on this reference tree based on all 100 of the best trees (Sukumaran and Holder, 2010). MEGA5 was used to create a condensed tree which collapsed all nodes with less than 70% bootstrap support (Tamura et al., 2011).

Recombination analysis

Recombination analysis was performed on all isolates using RDP4.22 in order to check the actual structures against the strain types as previously determined using a simple UPGMA. Six recombination analyses (RDP, GENECONV, Chimaera, MaxChi, Bootscan, and SiScan) were used to identify potential recombinants and parents (Martin and Rybicki, 2000; Padidam et al., 1999; Posada and Crandall, 2001; Smith, 1992; Salminem et al., 1995; Gibbs et al., 2000). Isolates were checked one strain at a time, based on the previous strain typing, in order to simplify the RDP output. At least one representative potential parent of each possible parental type was included each time (e.g. Oz and CW for O, Mont for Eu-N, etc).

Results and Conclusions

In the course of phylogenetic studies of five large sections of the PVY genomes coming from multiple virus recombinants (Evans et al., 2014), several inconsistencies were encountered that involved ten whole genomes for PVY isolates deposited in the GenBank database. These were isolates FrKV15, GBVC_PVY_23, GBVC_PVY_9, 9703-3, SD1, 1104, FZ10, 1107, E30, and N-Nysa (Table 3.2). When this phylogenetic study was performed, a preliminary UPGMA-based strain typing placed these ten sequences into well-defined lineages with known PVY recombinant types (see Table 3.2). However, when individual trees were examined for each of the five PVY genome sections (described in detail in Evans et al., 2014), these ten isolates (Table 3.2) were found in unusual lineages for sections 1, (nt 1-500), 2 (nt 500-2,390), 4 (nt 5,850-9,200), or 5 (nt 9,200-9,700). The unexpected placements for sequences of PVY isolates N-Nysa, E30, FZ10, and 1104 are represented on Fig. 3.1, as illustrations. Specifically, both N-Nysa and E30 were placed among NA-N sequences when

section 4 (nt 5,850-9,200) was analyzed, while both FZ10 and 1104 were placed among NE-11 sequences for this same section 4 (Fig. 3.1 A and B).

These unusual placements suggested potential novel recombinant structures for these ten sequences, and all ten were subjected to a more detailed study using programs from the RDP4.22 package. The unusual placement of each of these isolates for a specific section phylogeny, e.g. NA-N for E30 and N-Nysa in section 4, and NE-11 for FZ10 and 1104 in section 4, was used as a guide for selection of corresponding potential parental sequences in analysis by the RDP4 programs. Consequently, long segments of the NE-11 sequences were found in SD1, 1104, FZ10, and 1107 isolates, while long segments of NA-N sequences were found in FrKV15, GBVC_23, 9703-3, E30, and N-Nysa isolates (Fig. 3.2). The recombination events were detected in all ten sequences by all six programs from the RDP package (RDP, GENECONV, Chimaera, MaxChi, Bootscan, and SiScan), with very high confidence ($p < 0.0001$). Overall, nine novel, previously unreported PVY recombinant structures were determined from these ten PVY sequences listed in Table 3.2. These structures are presented in Fig. 3.2. Isolate GVBC_PVY_9 resembled a relatively common recombinant, NTNb, but had its first recombinant junction shifted considerably, relative to the ordinary NTNb – to nt 383 from nt 500 (Fig. 3.2; Table 3.3). Two new recombinants, FrKV15 and GVBC_PVY_23, also resembled common recombinants, N:O and N-Wi, respectively, although both had one of the recombinant junctions (RJ2) shifted considerably, relative to the common recombinants – to nt 2281 from nt 2390 (Fig. 3.2; Table 3.3). In these two cases, however, the entire N-sequence present in typical $PVY^{N:O}$ and $PVY^{N\text{-}Wi}$ genome was replaced by a NA-N sequence fragment in FrKV15 and GVBC_PVY_23 (see Fig. 3.2). Three novel recombinants, 9703-3, N-Nysa and E30, were found built of three parental sequences

representing O, N, and NA-N genomes. These three parental sequences were never found in a single PVY recombinant genome before. Four other novel recombinants, SD1, 1104, FZ10, and 1107, contained sequences from O, N, and NE-11 parents – these three parental sequences were found in a PVY recombinant only once, in the PVY^{E} genome, but in a different recombination pattern (Fig. 3.2; Galvino-Costa et al., 2012a).

Positions of recombinant breakpoints or recombinant junctions for each of these novel structures are summarized in Table 3.3. Interestingly, parents (PVY^{O} and PVY^{N}) and positions of recombinant breakpoints in the 5'-halves of seven of the ten analyzed genomes, 9703-3, SD1, 1104, FZ10, 1107, E30, and N-Nysa, are completely "normal", i.e. typical of more common recombinants like PVY^{N-Wi} or PVY^{NTN} (Fig. 3.2, Table 3.3). In two cases, FrKV15 and GBVC_23, where NA-N-sequence is found in place of N-sequence characteristic of "ordinary" N:O and N-Wi types, recombinant breakpoint positions differ from those in the ordinary N:O and N-Wi recombinants (Table 3.3). The 3'-halves of the seven novel recombinants, 9703-3, SD1, 1104, FZ10, 1107, E30, and N-Nysa, on the other hand, are unique or "unusual" both in terms of parents, which include either NE-11 or NA-N sequences, and in positions of most of the breakpoints (Fig. 3.2, Table 3.3). Six of the nine novel recombinant patterns involved a breakpoint in or very close to the CP cistron, potentially affecting the serotype of the corresponding PVY isolate (see Fig. 3.2). Fig. 3.3 shows an alignment of the sixty N-terminal amino acids of the CP for select PVY isolates representing main strains of PVY exhibiting O or N serotypes in comparison to the ten isolates analyzed here; two serotype-specific epitopes were demonstrated to be located between aa 1 to 38 (Nikolaeva et al., 2012). Based on the analysis of this alignment, isolates FrKV15, GVBC_23, 9703-3 and 1107 can be predicted to have O serotype, while isolates GVBC_9, FZ10, E30, N-

Nysa, and SD1 can be predicted to exhibit N serotype (see Fig. 3.3). Isolate 1104 is likely to have N serotype as well, however, predictions in this case are less certain due to sequence quality problem for this isolate resulting in an unresolved aa at the key position of one of the N-specific epitopes (Fig. 3.3); the second N-specific epitope is unambiguous, suggesting an N serotype for 1104 (Fig. 3.3).

Discussion

Evolution of positive strand RNA viruses is driven by accumulation of mutations, recombination, and reassortment (Simon and Bujarski, 1994; Roossinck, 2003; Nagy and Simon, 1997; Nagy, 2008; Gibbs and Ohshima, 2010). PVY in potato presents an interesting model to study evolution of a virus through recombination and mutations that create a large pool of genomes for selection in different hosts and environments and help the virus to survive and even prosper (Singh et al., 2008; Quenouille et al., 2013; Karasev and Gray, 2013). An impressive genetic diversity of PVY strains and genetic variants provides an opportunity to address the biological role of different genetic determinants in the PVY genome responsible for adaptation to new hosts, overcoming resistance genes, and leading to better fitness of a particular strain in a particular host and environment (Quenouille et al., 2013; Karasev and Gray, 2013). However, the same genetic diversity of PVY leads to difficulties in interpreting biological data, and to uncertainties in conclusions stemming from the difficulties in classifying PVY strains and genetic variants.

Currently, typing of PVY strains is based on several different criteria (Singh et al., 2008; Karasev and Gray, 2013). The genetic classification taking into account reactions to PVY isolates in potato indicators carrying *Ny*, *Nc*, and *Nz* genes is the most comprehensive, but time-consuming and limited in use due to the lack of biological information for most of

the recombinants of PVY. Molecular classification of strains and genetic variants of PVY is based on sequence information available for multiple isolates of PVY, and may include detection of polymorphisms around main recombinant breakpoints in PVY recombinants by RT-PCR (Glais et al., 2002; Lorenzen et al., 2006b; Chikh-Ali et al., 2013), or sequence analysis and subsequent phylogenetic and recombination analyses (Lorenzen et al., 2006a; Hu et al., 2009; Karasev et al., 2011; Ogawa et al., 2012; Galvino-Costa et al., 2012b). However, even with the whole genome sequence available for a particular isolate of PVY, when typing PVY sequences to strain, it is essential to perform recombinant analysis in addition to other analyses in order to be certain of the true structure of individual isolates. Novel recombination patterns can be and are easily missed and wrongly identified as being one of the more common structures, if appropriate tools and approaches are not used. In this study, ten PVY isolates were found to represent nine novel recombinant structures or patterns (Fig. 3.2), only when subjected to the phylogenetic analysis of individual genome sections. These same isolates would have been mistyped if only subjected to strain typing using cruder methods like BLAST search or UPGMA analysis.

It is important to remember that a correct identification of the PVY isolate recombinant structure, and hence correct typing of an isolate to strain, is not merely a scientific exercise but may have long-lasting biological and economical consequences. For example, isolate N-Nysa, studied in this work, was used by a Polish breeding program for selection of potato cultivars resistant to different PVY strains, under the assumption it represented a typical isolate of the PVY^{NTN} strain (Zimnoch-Guzowska et al., 2013). As was shown here, N-Nysa is actually an unusual recombinant (Figs. 3.1A, 3.2), with a completely different genome structure and, hence, apparently different biology in potato cultivars

carrying resistance genes; this biology has yet to be determined. Consequently, breeding data and conclusions obtained using this PVY isolate cannot be applied to the PVY^{NTN} strain resistance, and will likely need revision.

The nine novel recombinant structures found in PVY isolates (see Fig. 3.2) bring the total number of the recombinant patterns known for PVY up to 25. Five parental sequences were found in these 25 recombinants, representing non-recombinant PVY^{O}, PVY^{N}, PVY^{NA-N}, and PVY^{C}, and also a partial sequence of a recombinant NE-11 (Fig. 3.2). The new recombination patterns represented by isolates FrKV15 and GBVC_23 (Bahrami Kamangar et al., 2014) may pose a unique challenge for PVY strain detection and differentiation, since both RT-PCR typing methodologies currently in use (Lorenzen et al., 2006; Chikh-Ali et al., 2013) would identify these isolates as possible PVY^{O} based on the presence of one O-specific PCR band. Six of the nine novel recombination patterns found contain sequences from three different parents, $PVY^{O}/PVY^{N}/PVY^{NA-N}$ or PVY^{O}/PVY^{N}/PVY-NE11, and therefore represent products of multiple recombination events resulting in complex structures. The diversity of the PVY recombinants appears to be greater than previously suspected (Hu et al, 2009). Understanding the driving forces behind this extensive PVY evolution through recombination will require biological tests.

Acknowledgements

This project was funded in part through grants from the United States Department of Agriculture (USDA)-NIFA-NRI (2009-35600-05025), and from the Idaho Potato Commission. The authors thank Dr. Celeste Brown for technical assistance with RDP4.22.

References

Bahrami Kamangar, S., Smagghe1, G., Maes, M., and De Jonghe, K. (2014) *Potato virus Y* (PVY) strains in Belgian seed potatoes and first molecular detection of the N-Wi strain. Journal of Plant Diseases and Protection 121: 10–19.

Berger, P.H., Adams, M.J., Barnett, O.W., Brunt, A.A., Hammond, J., Hill, J.H., Jordan, R.L., Kashiwazaki, S., Rybicki, E., Spence, N., Stenger, D.C., Ohki, S.T., Uyeda, I., van Zaayen, A., Valkonen, J., Vetten, H.J., (2005). Family *Potyviridae*. Virus taxonomy, Classification, Nomenclature of Viruses. Eighth Report of the International Committee on Taxonomy of Viruses. Academic Press, pp. 819–841.

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

Chikh-Ali, M., Rowley, J.S., Kuhl, J.C., Gray, S.M., and 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. *American Journal of Potato Research*, in press.

Edgar, R.C., (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research 32(5), 1792-97.

Evans, K.J., Brown, C.J., Gray, S.M., Karasev, A.V., (2014). Phylogenetic study of the origins of recombinant strains of Potato Virus Y. In preparation.

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

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

Gibbs, A., Ohshima, K., (2010). Potyviruses and the digital revolution. Ann. Rev. Phytopathology, 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 16573–582.

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. *Archives of Virology* 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 Disease 94, 1384–97.

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.V., 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. Pot. Res. 87(1), 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. and Gray, S.M. (2013) Continuous and emerging challenges of *Potato virus Y* in potato. *Annual Review of Phytopathology* 51: 571-586.

Kerlan, C., Nikolaeva, O.V., Hu, X., Meacham, T., Gray, S.M.,and 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.

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(21), 2947-2948.

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.

Lorenzen, J. H., Meacham, T., Berger, P. H., Shiel, P. J., Crosslin, J. M., Hamm, P. B. & Kopp, H. (2006). Whole genome characterization of Potato virus Y isolates collected in the western USA and their comparison to isolates from Europe and Canada. Archives of Virology 151, 1055-1074.

Martin, D.P., Rybicki, E., (2000). RDP: detection of recombination amongst aligned sequences. Bioinformatics 16562–563.

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(5), 674-683.

Nagy, P. D. (2008). Recombination in Plant Viruses. In *Plant Virus Evolution* pp. 133-156. Edited by M. J. Roossinck: Springer Berlin Heidelberg.

Nagy, P. D. & Simon, A. E. (1997). New Insights into the Mechanisms of RNA Recombination. Virology 235, 1-9.

Nikolaeva, O.V., Roop, D., Galvino-Costa, S.F.B., Figueira, A.R., Gray, S.M., and Karasev, A.V. (2012) Epitope mapping for monoclonal antibodies recognizing tuber necrotic strains of Potato virus Y. American Journal of Potato Research 89: 121-128.

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. Journal of Phytopathology 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.

Quintero-Ferrer, A., Robles-Hernandez, L., Gonzalez-Franco A.C., Kerlan, C., and Karasev, A.V. (2014) Molecular and biological characterization of a recombinant isolate of Potato virus Y from Mexico. Archives of Virology, published on-line January 9, 2014 (DOI 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.

Roossinck, M. J. (2003). Plant RNA virus evolution. Current Opinion in Microbiology 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. Retroviruses 11, 1423–1425.

Schubert, J., Fomitcheva, V., Sztangret-Wiśniewska, J., (2007). Differentiation of Potato virus Y strains using improved sets of diagnostic PCR-primers. Journal of Virological Methods 140: 66–74.

Simon, A. E. & Bujarski, J. J. (1994). RNA-RNA Recombination and Evolution in Virus-Infected Plants. *Annual Review of* Phytopathology 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. University of Kansas Science Bulletin 38, 1409–1438.

Sukumaran, J., Holder, M.T., (2010). A Python library for phylogenetic computing. Bioinformatics 26, 1569-1571.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28, 2731-2739.

Zimnoch-Guzowska, E., Yin, Z., Chrzanowska, M., and Flis, B. (2013). Sources and effectiveness of potato PVY resistance in IHAR's breeding research. American Journal of Potato Research 90: 21-27.

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.

Supplementary Information

Table 3.1. Whole genome sequences used for analyses in this work. TBA = to be assigned.

GenBank Isolates			
	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	Unclassified
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	Unclassified
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	C
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-
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

Newly Sequenced Isolates					
	Isolate Name	**GenBank Accession**	**Strain**	**Serology**	**Multiplex**
1	AL100001	TBA	C	O	C
2	CA14	TBA	O	O	O
3	CO11	TBA	O5	O5	O
4	CO86	TBA	O	O	O
5	ID1_4_32B	TBA	O5	O5	O
6	ID_1_7_12B	TBA	O	O	O
7	ID_1258	TBA	O	O	O
8	ID1_1_3A	TBA	N-Wi	O	N-Wi
9	ID1_3_11B	TBA	N-Wi	O	N-Wi
10	ID11_13_11b	TBA	NTNa	N	NTNa
11	ID11_13_12A	TBA	N-Wi	O	N-Wi
12	ID12_102IC3	TBA	NTNa	N	NTNa
13	ID12_110Ban1	TBA	O5	O5	O
14	ID12_22RN8	TBA	NTNa	N	NTNa
15	ID12_401Chf	TBA	N-Wi	O	N-Wi
16	ID125	TBA	N-Wi	O	N-Wi
17	ID1280	TBA	NE-11	N	NE-11
18	ID13_148Oth	TBA	N-Wi	O	N-Wi
19	ID13_610Brw	TBA	NE-11	N	NE-11
20	ID21	TBA	NE-11	N	NE-11
21	ID26	TBA	NE-11	N	NE-11
22	ID281_O5	TBA	O5	O5	O
23	ID38	TBA	NTNa	N	NTNa
24	ID50	TBA	NTNa	N	NTN
25	ID89	TBA	N-Wi	O	N-Wi
26	ID90	TBA	N:O	O	N:O
27	Linda14	TBA	N-Wi	O	Uncl.
28	ME_236_120	TBA	O	O	O
29	ME_236_71	TBA	O	O	O

30	ME10	TBA	NTNa	N	NTNa
31	ME100004	TBA	N:O	O	N:O
32	ME100007	TBA	O	O	O
33	ME100008	TBA	O5	O5	O
34	ME100011	TBA	NTNa	N	NTNa
35	ME100031	TBA	NTNa	N	NTNa
36	ME110008	TBA	NTNa	N	NTNa
37	ME110032	TBA	NTNa	N	NTNa
38	ME200cornell	TBA	O5	O5	O
39	ME4	TBA	NTNa	N	NTNa
40	ME81	TBA	N:O	O	N:O
41	ME9	TBA	NTNa	N	NTNa
42	MI090004	TBA	N:O	O	N:O
43	MI110011	TBA	N-Wi	O	N-Wi
44	MN10c_26	TBA	N:O	O	N:O
45	MN121	TBA	N:O	O	N:O
46	MN13a_39	TBA	N:O	O	N:O
47	MN15_G_52	TBA	N-Wi	O	N-Wi
48	MN21	TBA	N-Wi	O	N-Wi
49	MN85	TBA	NTNa	N	NTNa
50	MSU_45-384a	TBA	Eu-N	N	Eu-N
51	MSU_59-383b	TBA	Eu-N	N	Eu-N
52	MT100006	TBA	Eu-N	N	Eu-N
53	MT100010	TBA	O5	O5	O
54	MT100017	TBA	Eu-N	N	Eu-N
55	MT29	TBA	O5- (no marker)	O	O
56	MT52	TBA	N:O	O	N:O
57	MT63	TBA	O5- (no marker)	O	O
58	ND100040	TBA	NE-11	N	NE-11
59	ND110037	TBA	N:O	O	N:O
60	ND121	TBA	N:O	O	N:O
61	ND18	TBA	N:O	O	N:O
62	ND2	TBA	N-Wi	O	N-Wi
63	ND23	TBA	N:O	O	N:O
64	ND35	TBA	O5	O	O
65	ND65	TBA	N:O	O	N:O
66	ND68	TBA	N:O	O	N:O
67	ND71	TBA	N:O	O	N:O
68	ND98	TBA	N-Wi	O	N-Wi
69	ND99	TBA	N:O	O	N:O
70	NE38	TBA	O	O	O

71	NE40	TBA	N:O	O	N:O
72	NE6	TBA	O	O	O
73	NY090004	TBA	N:O	O	N:O
74	NY090029	TBA	NTNa	N	NTN
75	NY090031	TBA	O	O	O
76	NY100001	TBA	O	O	O
77	NY100002	TBA	O	O	O
78	NY100003	TBA	O	O	O
79	NY100086	TBA	O	O	O
80	NY110001	TBA	Unclassified	O	O
81	NY120001	TBA	N-Wi	O	N-Wi
82	NY120002	TBA	N-Wi	O	N-Wi
83	NY51	TBA	N:O	O	N:O
84	OR16	TBA	N:O	O	N:O
85	OR2	TBA	O	O	O
86	OR20	TBA	O	O	O
87	OR3	TBA	N-Wi	O	N-Wi
88	OR35	TBA	NTNa	N	NTNa
89	Pondo4	TBA	261-4 like	O	Uncl.
90	SU2	TBA	NE-11	N	NE-11
91	WA316	TBA	NE-11	N	NE-11
92	WA70	TBA	N:O	O	N:O
93	WA9	TBA	O	O	O
94	WI120018	TBA	NE-11	N	NE-11
95	WI120092	TBA	O	O	Inconclusive
96	WI120127	TBA	N:O	O	N:O
97	WI3	TBA	O	O	O
98	WI3406	TBA	O	O	O
99	WI62	TBA	N:O	O	N:O
100	WY1	TBA	O5- (no marker)	O	O
101	CO_225	TBA	O5	O5	O
102	CO_28	TBA	O5	O5	O
103	CO12	TBA	O5	O5	O
104	CO238	TBA	O5	O5	O
105	CO254	TBA	O5	O5	O
106	CO275	TBA	O5	O5	O
107	CO32	TBA	O5	O5	O
108	CO39	TBA	O5	O5	O
109	CO53	TBA	O5	O5	O
110	CO55	TBA	O5	O5	O
111	CO6	TBA	O5	O5	O

112	ID_1005	TBA	O5	O5	O
113	ID_236	TBA	O5- (no marker)	O	O
114	ME_222_18	TBA	O5	O5	O
115	ME_250_106	TBA	O5	O5	O
116	ME_250_20	TBA	O5	O5	O
117	ME_323_34	TBA	O	O	O
118	ND127	TBA	O	O	O
119	T1	TBA	O (unusual)	O	O

Table 3.2. PVY whole genome sequences subjected to recombination analysis in this work.

Isolate name	GenBank accession	Origin	Original strain assignment*	UPGMA typing
FrKV15	HM991454	France	Atypical N-Wi	N:O
GVBC_PVY_23	JQ969040	Belgium	N-Wi	N-Wi
GVBC_PVY_9	JQ969037	Belgium	NTN	NTNb
9703-3	KC296432	China	-	N:O
SD1	EU182576	China	N	N-Wi
1104	KC296436	China	-	NTNb
FZ10	JN083842	China	Atypical NTN	NTNb
1107	KC296439	China	-	NTNb
E30	HM991453	Poland	Atypical NTN	NTNb
N-Nysa	FJ666337	Poland	N	NTNa

*Strain assignment as described by the authors who deposited the sequence, - marks no strain assignment in the GenBank description file.

Table 3.3. Summary of recombinant junction positions in genomes of novel recombinants.

Isolate name	Breakpoint position in the alignment
FrKV15	2281
GVBC_PVY_23	383, 2281
GVBC_PVY_9	383, 2390, 5850, 9200
9703-3	2390, 6844, 7200
SD1	2390, 6267, 7014, 8568, 9375
1104, FZ10	500, 2390, 5850
1107	500, 2390, 5850, 8557, 9097
E30	500, 2390, 5850, 9000
N-Nysa	2390, 5850, 9000

Figure legends

Fig. 3.1. Phylogenetic analysis of the PVY genome section 4 for select PVY isolates from the NA-N lineage (A), and from the NE-11/E lineage (B). Yellow highlights the PVY sequences analyzed in this work, and red highlights the PVY-T13 unusual recombinant sequence found previously (Ogawa et al., 2012).

Fig. 3.2. A schematic diagram of *Potato virus Y* (PVY) recombinant structures. The ruler on top represents the PVY genome (ca. 9.7-kb), individual cistrons are presented as rectangles with corresponding designations, P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa, NIb, and CP. Potential parental sequences are grouped on top, with different parents shaded differently. A group of most common recombinants is designated "Main recombinants" and fragments originating from different parents are shaded accordingly. A large group of rare PVY recombinants is presented at the lower half of the figure, with segments from different parents shaded accordingly. The ten sequences analyzed in this work are boxed. Isolate names are highlighted in yellow when sequences were found to be recombinant in this work, and red when found to be recombinant elsewhere.

Fig. 3.3. Multiple alignment of the N-terminal CP sequences (aa 1-60) for select PVY isolates from different strains exhibiting O and N serotypes. Brackets to the right designate PVY strains and serotypes. Rectangles below the alignment mark epitope positions for monoclonals supplied by Bioreba-N (Bn), SASA-N, and Neogen-N (Sn/Nn). Arrows above the alignment designate amino acid positions that differ between serotypes O and N, numbers show the amino acid position; solid arrows indicate positions with a clear-cut difference between O and N serotype, and empty arrows show the positions where the difference between O and N serotype is not clear-cut. Epitope positions and designations are from Nikolaeva et al. (2012).

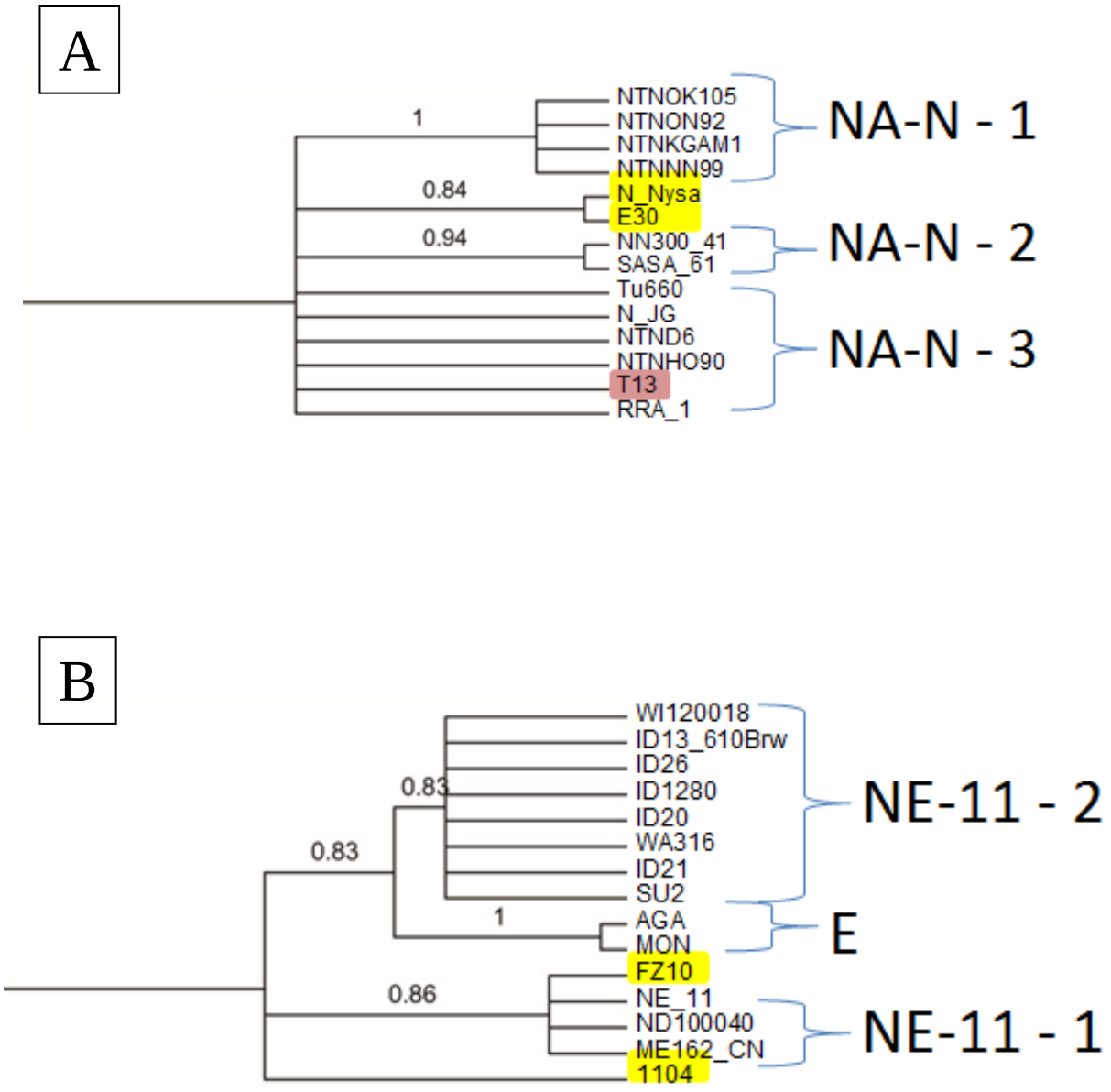

A
1
0.84
0.94
NTNOK105
NTNON92
NTNKGAM1
NTNNN99
N_Nysa
E30
NN300_41
SASA_61
Tu660
N_JG
NTND6
NTNHO90
T13
RRA_1
NA-N - 1
NA-N - 2
NA-N - 3
B
0.83
0.83
1
0.86
WI120018
ID13_610Brw
ID26
ID1280
ID20
WA316
ID21
SU2
AGA
MON
FZ10
NE_11
ND100040
ME162_CN
1104
NE-11 - 2
E
NE-11 - 1

Fig. 3.1

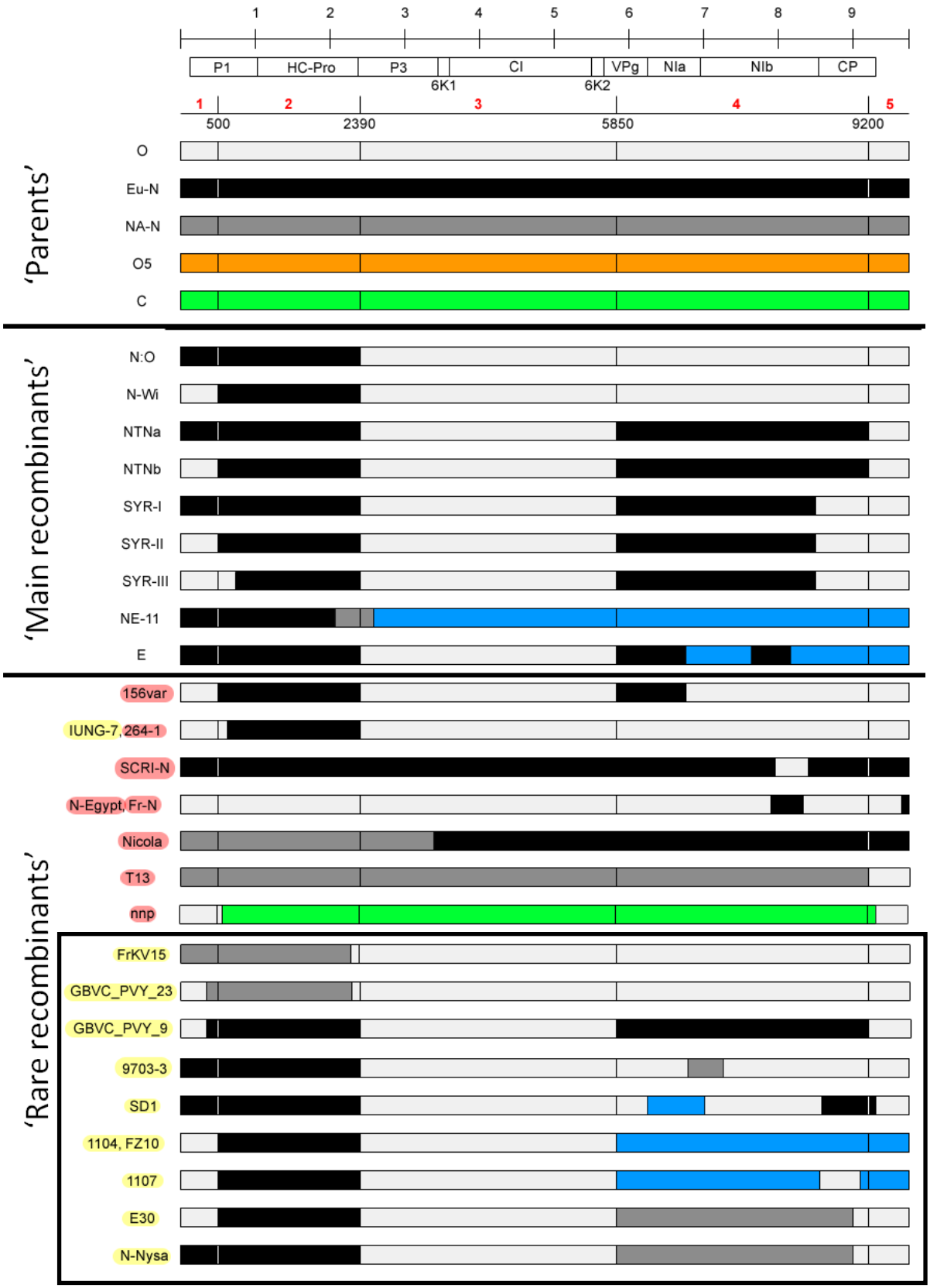

1
2
3
4
5
6
7
8
9
P1
HC-Pro
P3
6K1
CI
6K2
VPg
NIa
NIb
CP
1
2
3
4
5
500
2390
5850
9200
'Parents'
O
Eu-N
NA-N
O5
C
'Main recombinants'
N:O
N-Wi
NTNa
NTNb
SYR-I
SYR-II
SYR-III
NE-11
E
'Rare recombinants'
156var
IUNG-7, 264-1
SCRI-N
N-Egypt, Fr-N
Nicola
T13
nnp
FrKV15
GBVC_PVY_23
GBVC_PVY_9
9703-3
SD1
1104, FZ10
1107
E30
N-Nysa

Fig. 3.2

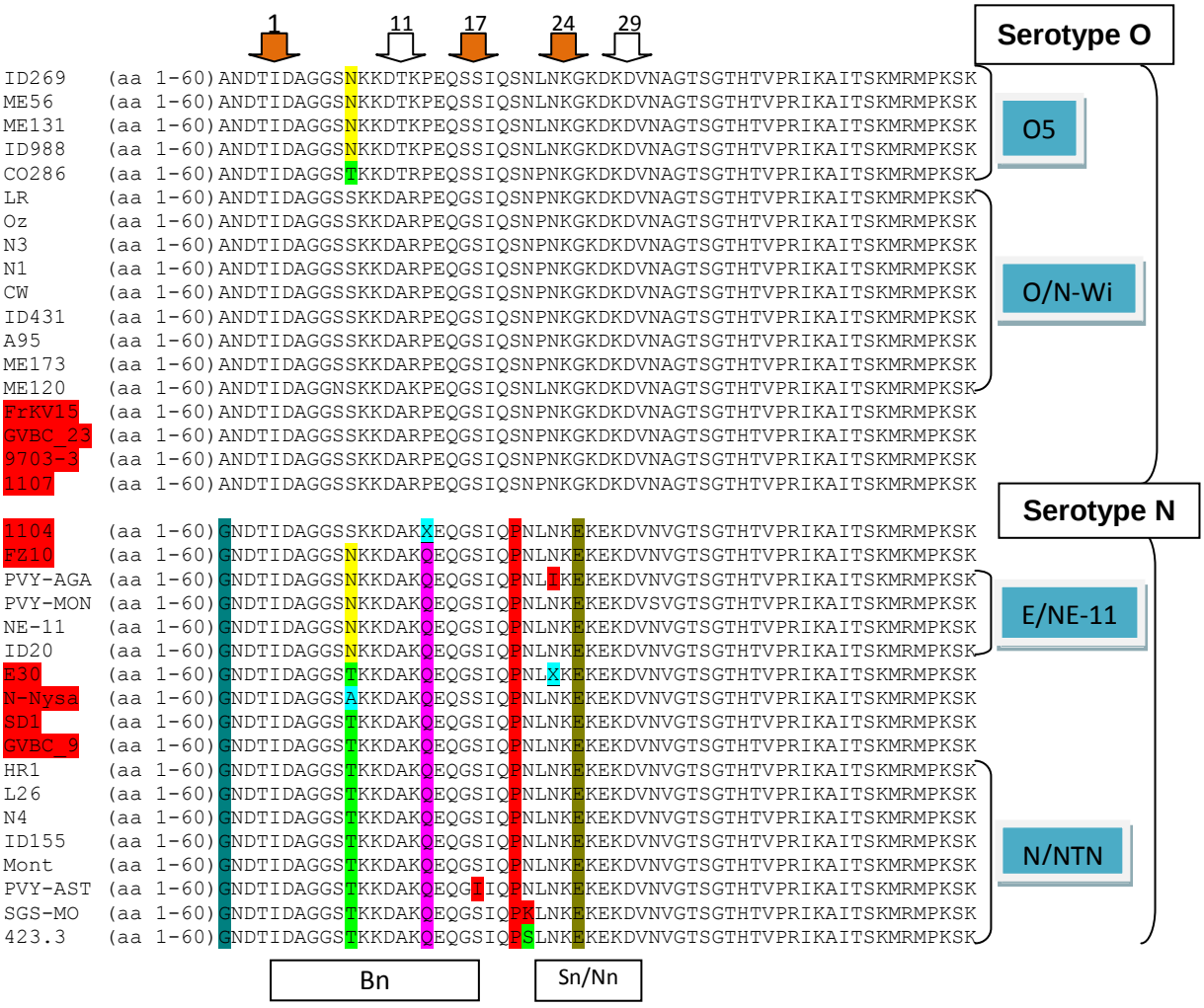

1
11
17
24
29
Serotype O
ID269 (aa 1-60)ANDTIDAGGSNKKDTKPEQSSIQSNLNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
ME56 (aa 1-60)ANDTIDAGGSNKKDTKPEQSSIQSNLNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
ME131 (aa 1-60)ANDTIDAGGSNKKDTKPEQSSIQSNLNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
ID988 (aa 1-60)ANDTIDAGGSNKKDTKPEQSSIQSNLNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
CO286 (aa 1-60)ANDTIDAGGSTKKDTRPEQSSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
O5
LR (aa 1-60)ANDTIDAGGSSKKDARPEQGSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
Oz (aa 1-60)ANDTIDAGGSSKKDARPEQGSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
N3 (aa 1-60)ANDTIDAGGSSKKDARPEQGSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
N1 (aa 1-60)ANDTIDAGGSSKKDARPEQGSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
CW (aa 1-60)ANDTIDAGGSSKKDARPEQGSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
ID431 (aa 1-60)ANDTIDAGGSSKKDARPEQGSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
A95 (aa 1-60)ANDTIDAGGSSKKDARPEQGSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
ME173 (aa 1-60)ANDTIDAGGSSKKDARPEQGSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
ME120 (aa 1-60)ANDTIDAGGNSKKDAKPEQGSIQSNLNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
O/N-Wi
FrKV15 (aa 1-60)ANDTIDAGGSSKKDARPEQGSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
GVBC_23 (aa 1-60)ANDTIDAGGSSKKDARPEQGSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
9703-3 (aa 1-60)ANDTIDAGGSSKKDARPEQGSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
1107 (aa 1-60)ANDTIDAGGSSKKDARPEQGSIQSNPNKGKDKDVNAGTSGTHTVPRIKAITSKMRMPKSK
Serotype N
1104 (aa 1-60)GNDTIDAGGSSKKDAKXEQGSIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
FZ10 (aa 1-60)GNDTIDAGGSNKKDAKQEQGSIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMKMPKSK
PVY-AGA (aa 1-60)GNDTIDAGGSNKKDAKQEQGSIQPNLIKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
PVY-MON (aa 1-60)GNDTIDAGGSNKKDAKQEQGSIQPNLNKEEKEKDVSVGTSGTHTVPRIKAITSKMRMPKSK
NE-11 (aa 1-60)GNDTIDAGGSNKKDAKQEQGSIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
ID20 (aa 1-60)GNDTIDAGGSNKKDAKQEQGSIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
E/NE-11
E30 (aa 1-60)GNDTIDAGGSTKKDAKQEQGSIQPNLXKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
N-Nysa (aa 1-60)GNDTIDAGGSAKKDAKQEQSSIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
SD1 (aa 1-60)GNDTIDAGGSTKKDAKQEQGSIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
GVBC_9 (aa 1-60)GNDTIDAGGSTKKDAKQEQGSIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
HR1 (aa 1-60)GNDTIDAGGSTKKDAKQEQGSIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
L26 (aa 1-60)GNDTIDAGGSTKKDAKQEQGSIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
N4 (aa 1-60)GNDTIDAGGSTKKDAKQEQGSIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
ID155 (aa 1-60)GNDTIDAGGSTKKDAKQEQGSIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
Mont (aa 1-60)GNDTIDAGGSTKKDAKQEQGSIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
PVY-AST (aa 1-60)GNDTIDAGGSTKKDAKQEQGTIQPNLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
SGS-MO (aa 1-60)GNDTIDAGGSTKKDAKQEQGSIQPKLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
423.3 (aa 1-60)GNDTIDAGGSTKKDAKQEQGSIQPSLNKEEKEKDVNVGTSGTHTVPRIKAITSKMRMPKSK
N/NTN
Bn
Sn/Nn

Fig. 3.3

CHAPTER 4: General Conclusions and Discussion

Conclusions and Discussion

Phylogenetic Analyses

Recombination is one of the primary means of evolution for potyviruses like PVY (Gibbs and Ohshima, 2010). It is well established that there are many recombinant structures for PVY (Glais et al., 2002; Lorenzen et al., 2006; Ogawa et al., 2008, 2012; Hu et al., 2009a,b; Karasev and Gray, 2013). The recombinant strains of PVY now dominate among PVY isolates circulating in potato in much of the world, including Europe, Africa, North America, and Brazil (Blanchard et al., 2008; Djilani-Khouadja et al., 2010; Gray et al., 2010; Galvino-Costa et al., 2012; Visser et al., 2012). However, phylogenetic reconstructions of the origins of PVY recombinants are complicated by the necessity to account for the recombination events.

Our approach to deal with this problem was to analyze the genome in five partial sequences between the four main recombinant junctions (Fig. 2.1). We also substantially expanded the number of PVY genomes analyzed, which included 119 newly sequenced genomes and 166 whole genomes from the GenBank database (Table 2.1).

Three main questions drove this section of the work: 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 recombinant events happen between PVY strains?

Examination of the phylogenetic trees presented in Figs. 2.2-2.4 suggested that there might be no monophyletic lineages of PVY recombinants (excepting clades too small to make any conclusions about monophyly or polyphyly, such as PVY^{E}, which only has two isolates).

Even the N:O lineage, comprising 30 or 31 corresponding sequences out of 32, was not monophyletic, with 2 (Fig. 2.2) or 1 (Fig. 2.3) additional clades comprising MI090004 and ND23 isolates. The answer to the second question may be easier, 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., 2012).

The answer to the third question is more complicated, however. On the one hand, only 12 novel recombinant types were found, representing just 13 isolates out of 285. On the other hand, the presence of certain recombinants in multiple clades 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 (Fig. 4.1) and also on the multiple clades characteristic of the same or similar recombinant types, we can conclude that recombination between different strains of PVY is relatively frequent.

Recombination Analyses

Currently, typing of PVY strains is based on several different criteria (Singh et al., 2008; Karasev and Gray, 2013). Molecular classification of strains of PVY is based on sequence information available for multiple isolates of PVY, and may include sequence analysis and subsequent phylogenetic and recombination analysis (Lorenzen et al., 2006a; Hu et al., 2009; Karasev et al., 2011; Ogawa et al., 2012; Galvino-Costa et al., 2012). However, even with the whole genome sequence available for a particular isolate, when typing PVY sequences to strain, it is crucial to perform recombinant analysis in addition to other analyses to be certain of the true structure. Novel recombination patterns can be, and are, easily missed and wrongly identified as being one of the more common structures when inappropriate tools and approaches are used. In this study, 13 PVY isolates were found to represent 12 novel

recombinant structures or patterns (Fig. 4.1) only when subjected to the phylogenetic analysis of individual genome sections. These same isolates would have been mistyped if only subjected to strain typing using crude methods like BLAST search or UPGMA analysis. Additionally, the proper parents of sections may not have been correctly identified if a representative of each potential parental strain had not been included in recombination analyses.

As a result of this thesis work, the total number of the recombinant patterns known for PVY is now brought up to 28. The diversity of the PVY recombinants appears to be far greater than previously concluded (Hu et al, 2009b), and now it needs to be subjected to biological tests in order to understand the driving forces behind this apparently extensive PVY evolution via recombination.

Future Directions

As is ever the case with science, proposing a question and getting the shape of an answer leads to even more questions and suggestions for future work. Continued sequencing of field samples, especially from foreign countries where recombinant diversity is greater, would broaden the scope of this work even further and add to the conclusions, perhaps bulking up the clades of some of the smaller sub-strains, or strains with very little representation such as PVY^{E} or the as of yet unnamed novel 1104/FZ10 recombinant structure.

The primer set developed and used in this thesis (Table 1.1) was for Sanger sequencing. Development of an NGS primer set for PVY would be an obvious next step. The author did work on this for several months, but it took two years of troubleshooting to get a good Sanger set and there was not time to properly troubleshoot an NGS set in addition – a

different set was needed due to different amplicon size requirements. However, the advantage would be the ability to sequence the entire population of PVY infecting a plant, not just the dominant or most easily sequenced strain. Naturally, it would also be a higher-throughput process.

A question that still needs answering is why does PVY recombine the way it does? More is now known about the possible structures than ever before, and patterns begin to emerge (e.g. section 4 is the most active for recombination, see Fig 4.1), but the question remains unanswered in full. It is likely that something is acting upon hairpin loops and A-T-rich regions, but it is as of yet unknown what those forces are or how they work (Hu et al., 2009b).

Ultimately, one of the major goals of studying PVY is to determine the incredibly complex relationship between potato cultivar, virus strain, and environment that causes PTNRD, and the development of ways to overcome it. This thesis represents an essential step along the long road that leads to this goal.

References

Blanchard, A., Rolland, M., Lacroix, C., Kerlan, C., and Jacquot, E., (2008). Potato virus Y: a century of evolution. Current Topics in Virology, 7: 21-32.

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. European J. of Plant Pathology. 126 : 4, 479-488.

Galvino-Costa, S.B., Figueira, A., Camargos, V.V., Geraldino, P.S., Hu, X., Nikolaeva, O.V., Kerlan, C., and Karasev, A.V. (2012) A novel type of *Potato virus Y* recombinant genome, determined for the genetic strain PVY^{E}. *Plant Pathology* 61: 388-398.

Gibbs, A., Ohshima, K. (2010). Potyviruses and the digital revolution. Ann. Rev. Phytopathology, 48: 205-223.

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. *Archives of Virology* 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 Disease 94, 1384–97.

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.

Karasev, A.V. and Gray, S.M. (2013) Continuous and emerging challenges of *Potato virus Y* in potato. *Annual Review of Phytopathology* 51: 571-586.

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. *Archives of Virology* 151, 1055-1074.

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 Research 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. Journal of Phytopathology 160: 661-673.

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.

Visser, J.C., Bellstedt, D.U., and Pirie, M.D. (2012). The Recent Recombinant Evolution of a Major Crop Pathogen, Potato virus Y. PLoS ONE 7(11): e50631. doi:10.1371/journal.pone.0050631.

Supplementary Information

Figure Legends

Fig. 4.1. A schematic diagram of all currently known *Potato virus Y* (PVY) recombinant structures, including those newly determined as part of this thesis. The ruler at the top represents the PVY genome (ca. 9.7-kb), individual cistrons are presented as rectangles below that with corresponding designations, P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa, NIb, and CP. Potential parental sequences are grouped above the top black horizontal dividing bar, with different parents colored differently. Recombinants are below the top black bar, and fragments originating from different parents are colored accordingly. The five major recombinant sections, based on common breakpoints, are designated in red numbers and the locations of breakpoints are given below that. Isolates highlighted in red are rare and previously published structures. Structures in the black box were newly determined in this study. Those highlighted in yellow are GenBank isolates whose novel structures are newly determined in this thesis. Those highlighted in green are novel structures which were sequenced from field isolates during the course of this study.

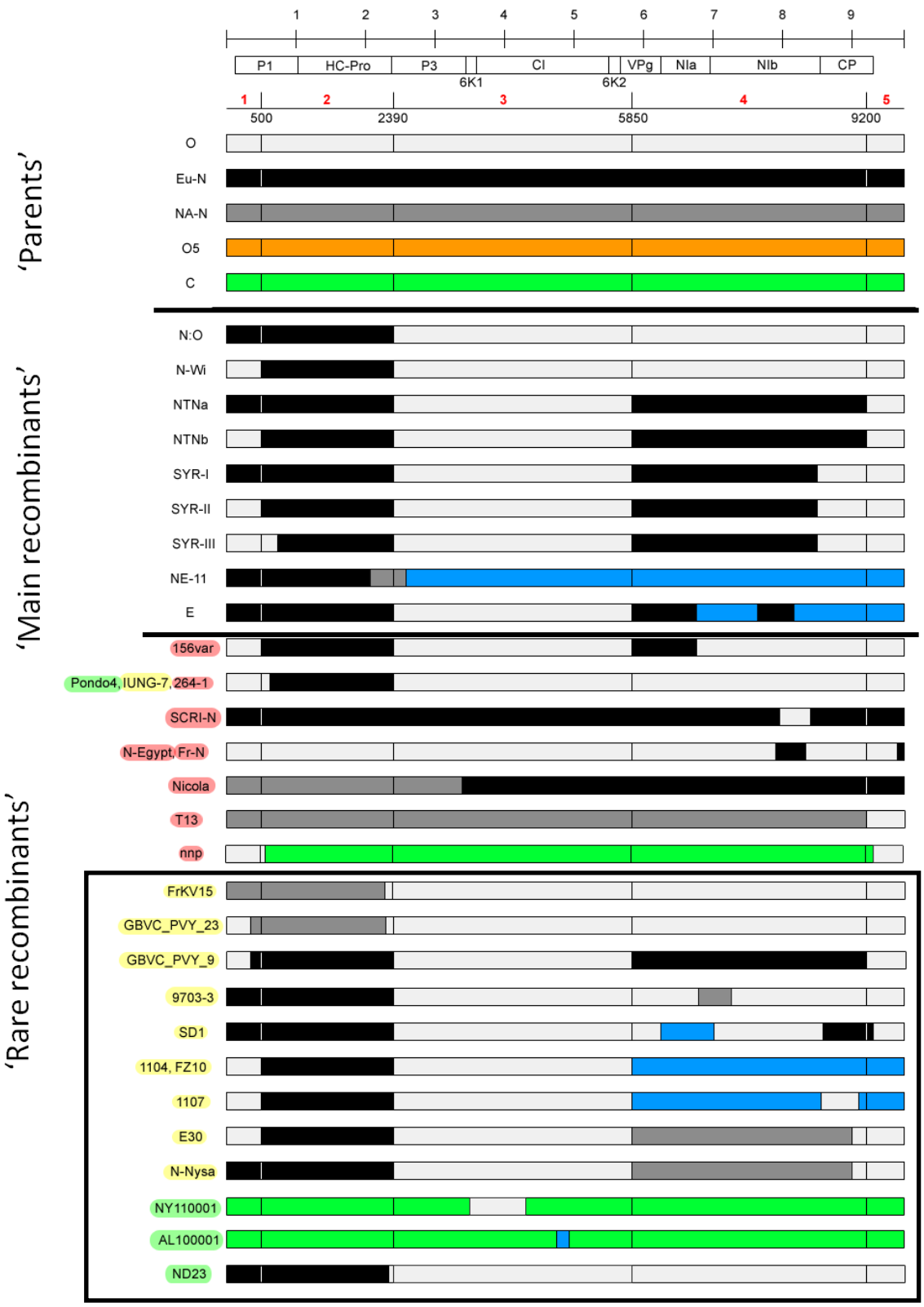

1
2
3
4
5
6
7
8
9
P1
HC-Pro
P3
6K1
CI
6K2
VPg
NIa
NIb
CP
1
2
3
4
5
500
2390
5850
9200
'Parents'
O
Eu-N
NA-N
O5
C
'Main recombinants'
N:O
N-Wi
NTNa
NTNb
SYR-I
SYR-II
SYR-III
NE-11
E
'Rare recombinants'
156var
Pondo4, IUNG-7, 264-1
SCRI-N
N-Egypt, Fr-N
Nicola
T13
nnp
FrKV15
GBVC_PVY_23
GBVC_PVY_9
9703-3
SD1
1104, FZ10
1107
E30
N-Nysa
NY110001
AL100001
ND23

Fig. 4.1