RESOURCE ANNOUNCEMENT

A Genome Resource for 192 *Verticillium dahliae* Isolates Infecting Potatoes Across Canada

Tanya Arseneault,[1] Joël Lafond-Lapalme,[1] Chary Esteban Quinche Gonzalez,[1] Katerina S. Jordan,[2] Dmytro Yevtushenko,[3] Ryan Barrett,[4] Hervé Van der Heyden,[5] Dahu Chen,[6] Mario Tenuta,[7] and Benjamin Mimee[1,†]

[1] Saint-Jean-sur-Richelieu Research and Development Centre, Agriculture and Agri-Food Canada, Saint-Jean-sur-Richelieu, QC, J3B 3E6, Canada
[2] Department of Plant Agriculture, University of Guelph, Guelph, ON, N1G 3B9, Canada
[3] Department of Biological Science, University of Lethbridge, Lethbridge, AB, T1K 3M4, Canada
[4] Prince Edward Island Potato Board, 90 Hillstrom Ave, Charlottetown, PE, C1E 2C6, Canada
[5] Compagnie de recherche Phytodata, Sherrington, QC, J0L 2N0, Canada
[6] Fredericton Research and Development Centre, Agriculture and Agri-Food Canada, Fredericton, NB, E3B 4Z7, Canada
[7] Department of Soil Science, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada

Funding
Financial support was provided by the Agriculture and Agri-Food Canada AgriScience Program (ASC-18/19 CHC).

Keywords
minimum spanning network, phylogeny, population genomics, variant calling, *Verticillium dahliae*, whole-genome sequencing

Abstract

Verticillium dahliae is an important soilborne pathogen causing Verticillium wilt. It is also the primary causal agent of potato early dying, a disease complex involving the root-lesion nematode. Here, we report the whole-genome sequencing of 192 isolates of *V. dahliae* originating from the major potato production areas across Canada. Our results yielded a resource of 277,010 genetic variations that will be useful for genetic analyses and revealed the presence of two major lineages, both present in all provinces but exhibiting differences in regional prevalence.

Genomic Resource Announcement

Verticillium is a genus of fungi from the Ascomycota, of which several species are known to cause diseases on a wide diversity of cultivated plants. The most notorious pathogenic species are *V. dahliae* and *V. albo-atrum*, together responsible for billions of dollars of crop losses worldwide as a result of the disease Verticillium wilt (Pegg and Brady 2002). *V. dahliae* is also the primary causal agent of potato early dying (PED), a major yield-limiting factor in all potato production areas (Rowe et al. 1987). PED results in premature vine senescence and can limit potato tuber yield by as much as 50%. This disease is often observed as a complex of pathogens/pests with other pathogenic fungi (e.g., *V. albo-atrum*, *Colletotrichum coccodes*) and the root-lesion nematode (*Pratylenchus penetrans*) (Powelson and Rowe 1993). The *Verticillium* species have clonal lineages or vegetative compatibility groups (VCGs) that exhibit different levels of aggressiveness and adaptation to host species (Puhalla and Hummel 1983). In North America, VCG2 and 4 (divided into 4A and 4B) are predominant in potato fields, and strains from the VCG4A were shown to be more virulent (Joaquim and Rowe 1991). Early work using AFLP markers has established that isolates from a VCG subgroup were genetically similar (Collado-Romero et al. 2006). It was also possible to differentiate strains from the subgroup 4A and 4B isolated from potatoes using RFLP markers (Dobinson et al. 2000). The publication of a reference genome for *V. dahliae* (Klosterman et al. 2011)

†Corresponding author: B. Mimee; benjamin.mimee@agr.gc.ca

e-Xtra: Supplementary material is available online.

The author(s) declare no conflict of interest.

Accepted for publication 31 March 2023.

has opened up new possibilities for population genetics and virulence studies. Genotyping-by-sequencing confirmed that isolates could be assigned to a clonal lineage using single-nucleotide polymorphisms (SNPs) (Milgroom et al. 2014; Rafiei et al. 2018). Comparative genomics with related species and functional genomics of genes encoding virulence factors have deepened our understanding of plant-pathogen interactions and responses to nutrient stress (Klimes et al. 2015). Still, many questions remain open on the genetic basis of host range, symptomatology, and sexual reproduction, as well as on the population structure and evolution of *Verticillium* species (Chen et al. 2021). In Canada, *V. dahliae* is present in all potato growing regions, but its diversity, and lineage composition, as well as the identity of interacting partners in PED, are still unknown. In this study, we sequenced the genome of 192 isolates of *V. dahliae* originating from the major potato production areas across Canada.

Soil sampling was carried out in commercial potato fields across Canada in the fall of 2019 and 2020. For each location, 40 soil cores 0 to 25 cm deep were collected using a 6-cm diameter Dutch auger following a "W" pattern across the field. *Verticillium* spp. were isolated using a baiting strategy by growing a Russet Burbank potato in 15-cm pots filled with a mix of 300 ml of the sampled soil and 300 ml of pasteurized soil (2:1 black earth/sand) in a growth chamber with a 16:8-h light/dark cycle and 23/18°C temperature. After 2 months, the stems of these plants were surface-sterilized, sliced, and transferred onto potato dextrose agar plates containing 0.01% of streptomycin sulfate for 2 weeks at 23°C. Then, colonies were observed under a stereomicroscope, and those showing conidiophores with verticillate branching were transferred to Sorenson's NP-10 semi-selective medium and incubated at 23°C, without light for 2 weeks to promote the production of microsclerotia. All the isolates with morphological characters compatible with *V. dahliae* (absence of yellow pigment, short conidia, and presence of microsclerotia), as described by Inderbitzin et al. (2011), were kept. Pure isogenic lines were obtained by plating of single conidia on potato dextrose agar. DNA was extracted from each sample with the Qiagen DNeasy PowerSoil Pro kit using 10 mg of freeze-dried mycelium. The extracted DNA was then sent to the Génome Québec Innovation Centre (Montreal, Canada) for library preparation using the NEBNext Ultra II DNA Library Prep Kit and for whole-genome sequencing on an Illumina NovaSeq 6000 S4 (PE150). Sequences were aligned to *V. dahliae* reference genome VdLs.17 (GCA_000150675.2) (Klosterman et al. 2011) with BWA v0.7.17 (Li and Durbin 2009), using default settings. Samtools v1.15 (Li et al. 2009) was then used to sort the respective sequence files. Variant calling was done with freebayes v1.3.6 (Garrison and Marth 2012) with the following parameters: –min-alternate-fraction 0.2, –use-best-n-alleles 4, –min-alternate-total 30, –min-alternate-count 5, –min-coverage 500. The resulting Variant Call Format (VCF) was used for the subsequent genetic analysis using the vcfR package (Knaus and Grünwald 2017). Phylogenetic relationships were observed in a Nei's genetic distance tree using the R packages poppr and ggtree (Kamvar et al. 2014; Yu 2020). Haplotype clustering for subgroup delimitation was assessed using the Bayesian information criterion to determine the best number of clusters (Jombart et al. 2010) to be included in a discriminant analysis with the adegenet R package (Jombart 2008). A minimum spanning network was created using this clustering information in poppr to visualize the relationships among individual isolates. We also compared our dataset with the data from Bautista-Jalón et al. (2021) to infer VCGs associated with our main genetic lineages. The two VCF files were merged with bcftools (Danecek et al. 2021) and filtered to keep only those in common. This was done only on the first chromosome (CP010980.1) to simplify the process. The resulting VCF was used to build a phylogenetic tree as above.

A total of 192 pure isogenic lines from six provinces were obtained (Fig. 1A; Supplementary Table S1). On average, 13.1 M sequencing reads were obtained per sample, corresponding to a genome coverage of 90×. Alignment to the *V. dahliae* reference genome VdLs.17 revealed a total of 277,010 genetic variations distributed at 259,497 loci and including 207,606 SNPs, 13,112 deletions, 11,857 insertions, and 28,873 complex mutations (Supplementary Table S2). A phylogenetic tree indicated the presence of two clades containing 74 and 118 isolates (Fig. 1B). Each of these lineages was present in all provinces, but not in the same proportions, with the first lineage dominating in Western Canada and the second lineage more prevalent in Eastern Canada (Fig. 1C). A comparison with the SNP dataset obtained on phenotyped (VCG) isolates by Bautista-Jalón et al. (2021) indicated that the first lineage corresponds to the VCG4A, whereas the second lineage is VCG4B (Supplementary Fig. S1). Among each lineage, several haplotypes were observed and clustered into three main genetic groups for lineage 4A and six groups for lineage 4B according to Bayesian information

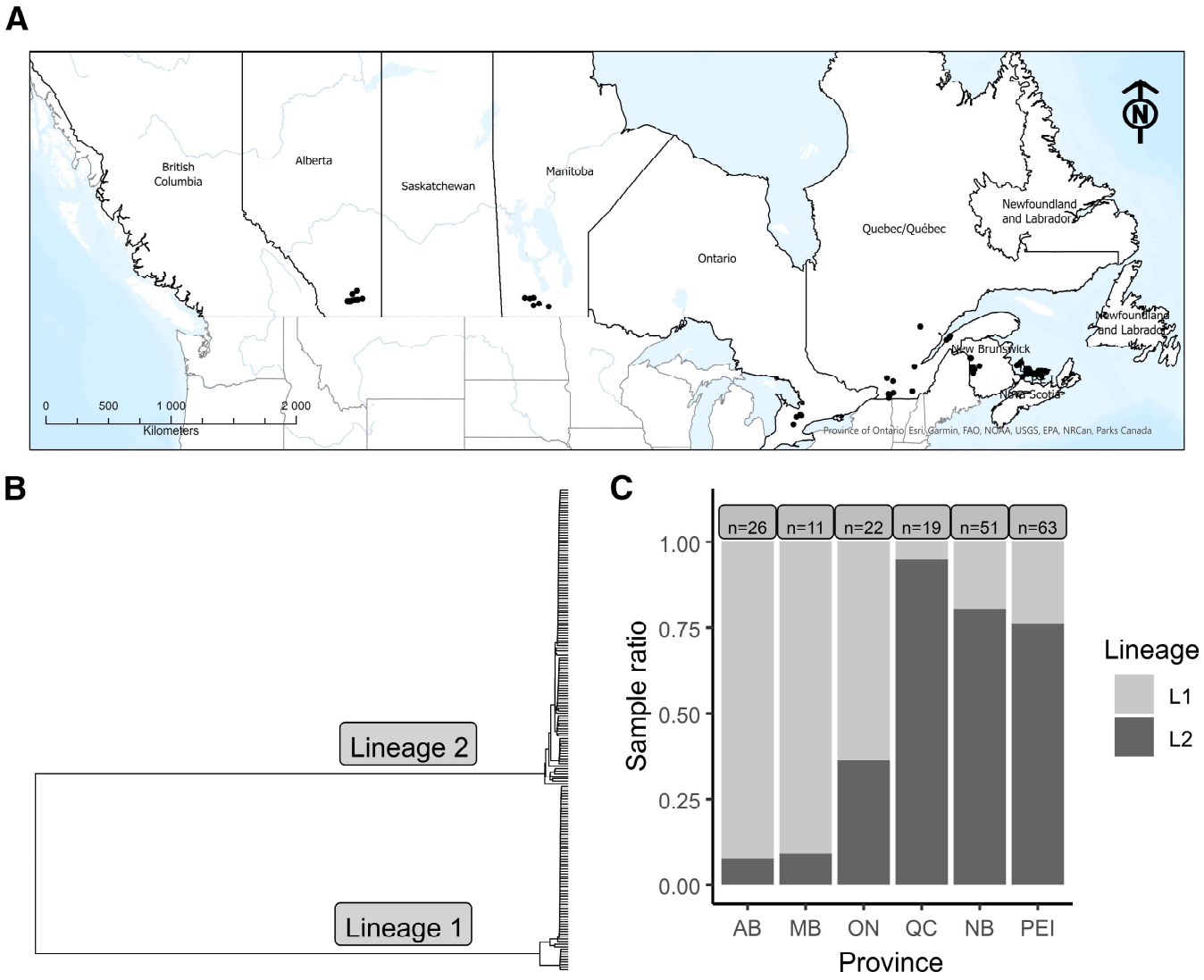

Fig. 1. Geographical origin, phylogenetic tree and distribution of the main *Verticillium dahliae* lineages in Canada. **A,** Map of Canada showing the sampling locations (black dots). **B,** Phylogenetic tree based on genetic distance (Nei's) between isolates of *V. dahliae*; the tree shows the evolutionary relationships between isolates based on the similarities in the single-nucleotide polymorphism dataset from the whole genome. **C,** Proportion of the isolates from each of the main lineages in each province; "n" is the total number of isolates analyzed for each province.

criterion analysis (Supplementary Fig. S2). A minimum spanning network revealed that the distribution of these haplotypes was spread across Canada without any regional structure (Fig. 2). Multiple haplotypes were found in a single field (Supplementary Table S1). These genomics resources are important to determine the genetic structure of *V. dahliae* populations and for comparative genomics. This will enable population genomics studies to identify genome features that are involved in pathogenicity and the development of new diagnostics applications to improve disease management.

Data Availability

The whole-genome sequencing project has been deposited at GenBank under the BioProject number PRJNA890653. The accession numbers of individual isolates are SRR21907825 to SRR21908016. Scripts and links to resources can be found at https://github.com/Mimee-Lab/canpednet_Vdahliae.

Acknowledgments

We thank Josiane Vaillancourt for fungal isolation and culturing; Vanessa Currie, Kamrun Nahar, Sebastian Ibarra, Eileen Beaton, Morgan McNeil, Atta Ur Rahman, Terri Fairman, Andréanne Sauvageau, Samuel Morissette, Samuel Gagnon, and André Gagnon

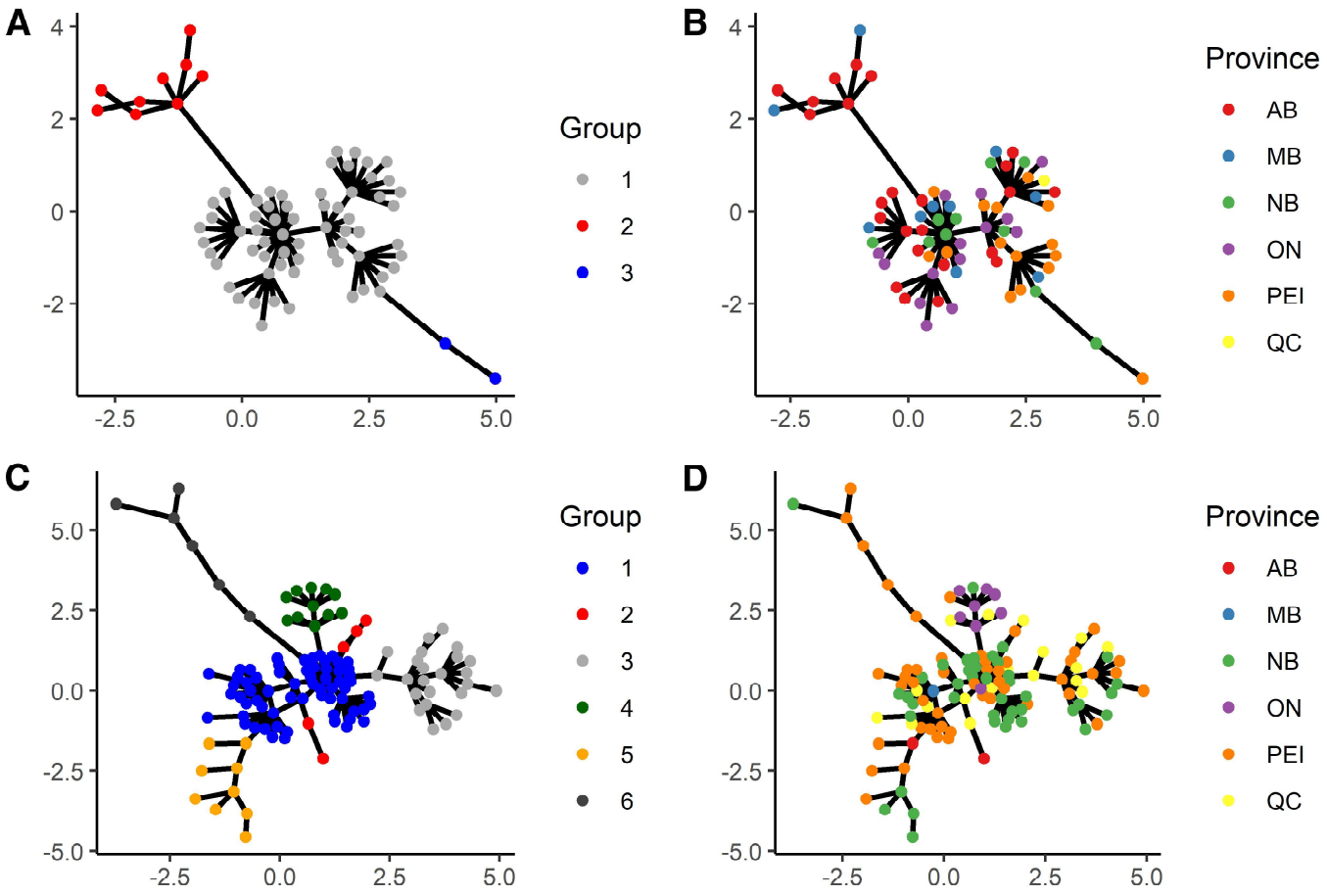

Fig. 2. Minimum spanning network of *Verticillium dahliae* isolates from Canada with genetic clusters delineated using the Bayesian information criterion. **A,** Lineage 1 (VCG4A) with colors indicating genetic clusters; **B,** lineage 1 (VCG4A) with colors indicating provinces of origin; **C,** lineage 2 (VCG4B) with colors indicating genetic clusters; and **D,** lineage 2 (VCG4B) with colors indicating provinces of origin.

for soil sampling; Philippe Vigneault for geomatics support; and Pierre-Yves Véronneau for technical assistance and bioinformatics.

Literature Cited

Bautista-Jalón, L. S., Frenkel, O., Tsror (Lahkim), L., Malcolm, G. M., Gugino, B. K., Lebiush, S., Hazanovsky, M., Milgroom, M. G., and Jiménez-Gasco, M. M. 2021. Genetic differentiation of *Verticillium dahliae* populations recovered from symptomatic and asymptomatic hosts. Phytopathology 111:149-159.

Chen, J.-Y., Klosterman, S. J., Hu, X.-P., Dai, X.-F., and Subbarao, K. V. 2021. Key insights and research prospects at the dawn of the population genomics era for *Verticillium dahliae*. Annu. Rev. Phytopathol. 59:31-51.

Collado-Romero, M., Mercado-Blanco, J., Olivares-García, C., Valverde-Corredor, A., and Jiménez-Díaz, R. M. 2006. Molecular variability within and among *Verticillium dahliae* vegetative compatibility groups determined by fluorescent amplified fragment length polymorphism and polymerase chain reaction markers. Phytopathology 96:485-495.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., and Li, H. 2021. Twelve years of SAMtools and BCFtools. GigaScience 10:giab008.

Dobinson, K. F., Harrington, M. A., Omer, M., and Rowe, R. C. 2000. Molecular characterization of vegetative compatibility group 4A and 4B isolates of *Verticillium dahliae* associated with potato early dying. Plant Dis. 84:1241-1245.

Garrison, E., and Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. http://arxiv.org/abs/1207.3907 (accessed January 23, 2023).

Inderbitzin, P., Bostock, R. M., Davis, R. M., Usami, T., Platt, H. W., and Subbarao, K. V. 2011. Phylogenetics and taxonomy of the fungal vascular wilt pathogen *Verticillium*, with the descriptions of five new species. PLoS One 6:e28341.

Joaquim, T. R., and Rowe, R. C. 1991. Vegetative compatibility and virulence of strains of *Verticillium dahliae* from soil and potato plants. Phytopathology 81:552-558.

Jombart, T. 2008. adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

Jombart, T., Devillard, S., and Balloux, F. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genet. 11:94.

Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. 2014. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2:e281.

Klimes, A., Dobinson, K. F., Thomma, B. P. H. J., and Klosterman, S. J. 2015. Genomics spurs rapid advances in our understanding of the biology of vascular wilt pathogens in the genus *Verticillium*. Annu. Rev. Phytopathol. 53:181-198.

Klosterman, S. J., Subbarao, K. V., Kang, S., Veronese, P., Gold, S. E., Thomma, B. P. H. J., Chen, Z., Henrissat, B., Lee, Y.-H., Park, J., Garcia-Pedrajas, M. D., Barbara, D. J., Anchieta, A., de Jonge, R., Santhanam, P., Maruthachalam, K., Atallah, Z., Amyotte, S. G., Paz, Z., Inderbitzin, P., Hayes, R. J., Heiman, D. I., Young, S., Zeng, Q., Engels, R., Galagan, J., Cuomo, C. A., Dobinson, K. F., and Ma, L.-J. 2011. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLoS Pathog. 7:e1002137.

Knaus, B. J., and Grünwald, N. J. 2017. vcfr: A package to manipulate and visualize variant call format data in R. Mol. Ecol. Resour. 17:44-53.

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

Milgroom, M. G., Jiménez-Gasco, M. M., García, C. O., Drott, M. T., and Jiménez-Díaz, R. M. 2014. Recombination between clonal lineages of the asexual fungus *Verticillium dahliae* detected by genotyping by sequencing. PLoS One 9:e106740.

Pegg, G. F., and Brady, B. L. 2002. Verticillium Wilts. CABI, New York, NY, U.S.A.

Powelson, M. L., and Rowe, R. C. 1993. Biology and management of early dying of potatoes. Annu. Rev. Phytopathol. 31:111-126.

Puhalla, J. E., and Hummel, M. 1983. Vegetative compatibility groups within *Verticillium dahliae*. Phytopathology 73:1305-1308.

Rafiei, V., Banihashemi, Z., Bautista-Jalon, L. S., del Mar Jiménez-Gasco, M., Turgeon, B. G., and Milgroom, M. G. 2018. Population genetics of *Verticillium dahliae* in Iran based on microsatellite and single nucleotide polymorphism markers. Phytopathology 108:780-788.

Rowe, R. C., Davis, J. R., Powelson, M. L., and Rouse, D. I. 1987. Potato early dying: Causal agents and management strategies. Plant Dis. 71:482-489.

Yu, G. 2020. Using ggtree to visualize data on tree-like structures. Curr. Protoc. Bioinformatics 69:e96.