Invasive exotic pathogens pose a threat to trees and forest ecosystems worldwide, hampering the provision of essential ecosystem services such as carbon sequestration and water purification. Hybridization is a major evolutionary force that can drive the emergence of pathogens. *Phytophthora ramorum*, an emergent pathogen that causes the sudden oak and larch death, spreads as reproductively isolated divergent clonal lineages. We use a genomic biosurveillance approach by sequencing genomes of *P. ramorum* from survey and inspection samples and report the discovery of variants of *P. ramorum* that are the result of hybridization via sexual recombination between North American and European lineages. We show that these hybrids are viable, can infect a host and produce spores for long-term survival and propagation. Genome sequencing revealed genotypic combinations at 54,515 single nucleotide polymorphism loci not present in parental lineages. More than 6,000 of those genotypes are predicted to have a functional impact in genes associated with host infection, including effectors, carbohydrate-active enzymes and proteases. We also observed post-meiotic mitotic recombination that could generate additional genotypic and phenotypic variation and contribute to homoploid hybrid speciation. Our study highlights the importance of plant pathogen biosurveillance to detect variants, including hybrids, and inform management and control.
Emerging infectious diseases caused by filamentous pathogens pose a threat to animals, crops and forest ecosystems worldwide. Globalization is driving the spread of pathogens into naive host populations and biotic homogenization provides an advantage to generalists. Disease outbreaks can have devastating ecological, economic and social impacts; they can reduce biodiversity, displace native species and threaten food security. Pests and pathogens can affect landscape resilience and reduce the ability of forests to provide essential ecosystem services such as carbon sequestration and water purification.

Hybridization can favor the emergence of infectious diseases by generating novel pathogen lineages with increased fitness that can colonize new niches. The large-scale genomic changes that result from hybridization generate new genotypic and phenotypic combinations upon which natural selection can act to modify traits such as pathogenicity and transmission. Selective forces can create the conditions that benefit the emergence of lineages with distinct traits that can result in lifestyle or host shifts. For example, human pathogens such as opportunistic yeasts and the malaria parasite have emerged following hybridization. Plant pathogens have also emerged following hybridization events that facilitated host jumps, increased virulence and host range expansion.

*Phytophthora ramorum* (Oomycota) is an exotic plant pathogen that emerged in the 1990s and is regulated and targeted for eradication and quarantines in Europe and North America. It can attack over 125 plant species and is responsible for the sudden oak death in California and Oregon and sudden larch death in the UK and France. The pathogen comprises clonal lineages that are reproductively isolated and diverged between 0.75 and 1.25 million years ago (MYA). Lineages NA1 and NA2 are confined to North America and lineage EU2 is only found in western Europe. Lineage EU1 is broadly distributed in Europe and the West Coast of North America. Eight additional lineages were recently described in Asia. Recombination can only take place between individuals of opposite mating types and was demonstrated in the laboratory between lineages EU1 (mating type A1) and NA1 (mating type A2) but the progeny displayed aberrant genotypic and phenotypic variation.

Preventing the emergence of new pathogenic requires the timely acquisition of data that can inform mitigation actions. Increasingly, pathogen genome sequencing is used within a biosurveillance framework to generate population genomic data across geographic and temporal scales that can be used to identify variants, recombinants and hybrids and inform pathway and risk analyses. Because of the potential for *P. ramorum* inter-lineage hybridization in western North America where lineages of opposite mating types co-occur, we analyzed the genomes of samples from Canada, the USA and Europe to characterize the patterns of variation and identified hybrids among two clonal lineages of the pathogen.

### Results

**Hybridization between divergent clonal lineages.** We discovered hybrids between the EU1 (mating type A1) and NA2 (mating type A2) clonal lineages of *P. ramorum* from a single nursery in British Columbia, Canada, where these lineages had been previously reported. The qPCR lineage genotyping pattern of isolates 16-237-021 and 16-284-032, collected from infected rhododendron plants, do not match the pattern of the known clonal *P. ramorum* lineages in North America and Europe (Supplementary Table 1). We analyzed the genomes of 95 *P. ramorum* isolates (Supplementary Table 2) from Europe and North America and mapped the sequence reads onto the reference genome, yielding 450,656 single nucleotide polymorphisms (SNPs). Of the 31,047 SNPs that were homozygous for different alleles in the EU1 and NA2 lineages, 96.6% were heterozygous in isolates 16-237-021 and 16-284-032 (Table 1). Genome-wide heterozygosity was nearly as high in the hybrid (*Hw* = 34.59%) as in the EU1 lineage (*Hw* = 35.22%; Supplementary Table 3). All other lineages, including the other putative parent of the hybrid, had lower heterozygosity (Table 1). A principal component analysis of all 450,656 SNPs placed the two putative hybrids between the EU1 and NA2 lineages and ancestry estimation assigned them with equal probability to those putative parental lineages (Fig. 1a, b). A phylogenetic network analysis placed the hybrids in a branch with shared reticulations with lineages EU1 and NA2, the pattern expected for recombination (Fig. 1c).

**First-generation homoploid hybrids generated via sexual recombination.** The hybrids are likely homoploid (without a change in chromosome copy number; Supplementary Fig. 1) and result from a first-generation (F1) recombination event. Analysis of phased genomic regions revealed distinct haplotypes at nuclear loci in the hybrids clustering either with a NA2 or an EU1 haplotype (Fig. 2a–c). The contribution of the putative parents to the nuclear genome of the hybrids was equal (Z-score = 8.78 ± 0.24; *p* < 0.0001; Supplementary Table 4) and they were assigned to a simulated EU1 x NA2 population with ≥ 0.999 probability (assignment to all other populations < 0.0001; Supplementary Fig. 2a). A neighbor-joining tree clustered the hybrid samples with simulated first-generation hybrids but not with simulated populations backcrossed to either parental lineage (Supplementary Fig. 3). The two hybrid samples share genotypes at 99.8% of the SNP loci and are thus likely clones derived from a single hybridization event, via sexual

### Table 1 Genotypic combinations in *Phytophthora ramorum* parental lineages and hybrid.

| Parental lineage genotypes | Hybrid genotype | Observed genotypes | Expected genotypes (sexual recombination) | Expected genotypes (heterokaryosis) |
|---------------------------|-----------------|--------------------|-------------------------------------------|-----------------------------------|
| 0/0 × 1/1                 | 0/0 or 1/1      | 1,049 (3.4%)       | 0 (0.0%)b                                 | 0 (0.0%)b                         |
| 0/1 x 0/1                | 0/1a            | 29,998 (96.6%)     | 31,047 (100.0%)b                          | 31,047 (100.0%)b                  |
| 0/2a                      | 12,295 (24.7%)  | 12,465 (25.0%)b    |                                           | 0 (0%)c                           |
| 1/2a                      | 12,222 (24.5%)  | 12,465 (25.0%)b    |                                           | 0 (0%)c                           |
| 0/1                       | 25,343 (50.8%)  | 24,930 (50.0%)b    |                                           | 49,860 (100%)c                    |
| 0/1 x (0/0 or 1/1)       | 0/0 or 1/1      | 114,604 (50.1%)    | 114,510 (50.0%)b                          | 114,510 (50.0%)b                  |
|                           | 0/11            | 114,415 (49.9%)    | 114,510 (50.0%)b                          | 49,860 (100%)c                    |

*a*Observed percentage of SNPs in each genotypic combination in the parental lineages and the *P. ramorum* hybrid samples 16-237-021 and 16-284-032 and expected genotypic combinations with sexual recombination and heterokaryosis following hyphal fusion (anastomosis).

*b*Genotypic combinations in the hybrid not observed in the parental lineages, totaling 54,515 SNPs.

*c*Observed genotype frequencies are not different from expected under the recombination model (chi-square test performed on 1000 subsamples of 100 random loci; *p* ≥ 0.05 for ≥ 95.0% of the subsamples).

*d*Observed genotype frequencies are different from expected under the recombination model (chi-square test performed on 1000 subsamples of 100 random loci; *p* ≤ 0.05 for 100% of the subsamples).
recombination, followed by clonal propagation. The observed genotypic combinations in the hybrids and the parental lineages support the hypothesis of sexual recombination (except for some excess of homozygosity observed in the hybrids [Table 1; see below]) but not of heterokaryosis following hyphal fusion (anastomosis; Table 1). The NA2 lineage likely acted as the “female” parent. We mapped the sequence reads of all 95 sequenced genomes to the mitochondrial genome of *P. ramorum* and retrieved their mitochondrial haplotypes (mitotypes). There are 103 polymorphic sites in the mitochondrial genomes of the four lineages of *P. ramorum*. The mitotypes of the EU1 and NA2 lineages differ at 42 positions. The NA2 and hybrid *P. ramorum* mitotypes are identical (with the exception of two SNPs in sample 16_0284_0032; Fig. 2d). This pattern indicates uniparental transmission of the mitochondrial genome.

**Predicted functional impact of hybridization.** The hybridization between two *P. ramorum* lineages that diverged approximately 1 MYA could impact fitness. These lineages differ in several traits, including growth, sporulation and aggressiveness during host infection. The growth rate of the hybrid isolates was intermediate between the parental lineages but significantly higher than the EU1 parent ($p = 3.75e-5 \times 10^{-5}$; Fig. 3d). The hybrids can produce both chlamydospores and sporangia (Fig. 3a, b, c), spores that are important for survival and spread. The hybrids were infectious on rhododendron, a common host, and caused lesions with sizes that overlapped with those of the other lineages (Fig. 3e). The hybrids have genotypes not previously observed at 54,515 SNPs that are either homozygous for different alleles in the parental lineages but heterozygous in the hybrid, or heterozygous in both parental lineages but homozygous in the hybrid (Table 1). Of those SNPs, 6,736 are non-synonymous mutations and 6,752 are predicted to have a moderate (e.g., non-synonymous mutations) to high (e.g., stop codons, frameshifts) impact (Supplementary Tables 5 and 6). Several of the genes that have novel genotypic combinations in the hybrids are associated with pathogenicity, including 51 RxLR and crinkler-like effectors, 79 carbohydrate active enzymes (CAZymes), and 11 peptidases (Supplementary Table 7).

**Mitotic recombination generates additional variation.** Mitotic recombination in the hybrid could provide additional genotypic and phenotypic variation in the pathogen population. We found 1,049 SNPs that were homozygous for different alleles in the parental lineages and also homozygous for one of those alleles in the hybrids, an unexpected pattern for sexual recombinants (Table 1). We observed that 90% of those SNPs were distributed in stretches of 2 to 82 (median = 4.0) contiguous homozygous positions (runs of homozygosity; ROH) in the hybrid (Supplementary Fig. 4), indicating that mitotic recombination had occurred in these regions. These ROH were in gene-poor regions enriched in transposable elements where RxLR effectors and putative avirulence factors are present (Supplementary Table 8).
Discussion

The discovery of hybridization among lineages of *P. ramorum* should cause concerns in the plant and forest health communities. Hybridization provides a source of genetic variation upon which natural selection can act to modify traits such as pathogenicity and transmission. Selective forces could create the conditions for homoploid hybrid speciation and the emergence of lineages with distinct traits. Neither hybridization nor recombination have previously been reported in the invasive range of *P. ramorum*. Previous studies that reported artificial crosses between *P. ramorum* EU1 and NA1 lineages produced progeny with reduced viability but that exhibited a broad range of pathogenicity. This could be explained by transgressive trait variation, a common outcome of hybridization in plants, where hybrid phenotypes exceed those of the parents, contributing to ecological niche divergence.

We show that the EU1 and NA2 lineages can form stable and viable hybrids in nature that can sporulate and infect a plant host. The growth phenotypes we observed in the hybrids were intermediate compared to the parental phenotypes. However, we have only conducted inoculations on a single host, rhododendron. Our genome analyses indicate that hybridization has a potential functional impact, with thousands of amino acid changes, several predicted to have a large effect. Some of the genes affected are associated with host infection and include effectors, known to be involved in pathogen–host interactions. A single amino acid polymorphism in a pathogen effector was recently shown to expand its binding spectrum, allowing the pathogen to adapt to new hosts, and demonstrating the advantage that novel mutations can confer. The genomic changes we report in the *P. ramorum* hybrids have the potential to impact diversification and adaptation.

The presence of sexual recombination among divergent *P. ramorum* lineages could modify the epidemiologic profile of the disease. Oospores, the spores produced following sexual reproduction, can survive adverse conditions such as drought and freeze and remain viable in the soil for years, making it possible for the pathogen to survive in the absence of a host thereby increasing the risk of accidental propagation. In the potato famine pathogen, *Phytophthora infestans*, sexual recombination became possible when lineages of the opposite mating type migrated from the center of origin and contributed to the emergence of novel virulence combinations, coinciding with the appearance of lineages that were resistant to fungicides and outbreaks that started earlier in the season, generating an increased disease risk.

In addition to meiotic recombination, we also observed mitotic recombination in the hybrids. Mitotic recombination was shown to accelerate adaptation in fungi and appears to be an important mechanism fueling evolution in *P. ramorum*, producing ROH in all lineages. Virulence differences and adaptation to environmental changes such as exposure to oxidative or heat stress and antifungal drugs have been associated with ROH in fungi and could increase genotypic and
phenotypic diversity and divergence between the *P. ramorum* lineages and the hybrid. So far, the *P. ramorum* EU1 × NA2 hybrid was only found in a single nursery in British Columbia, where the pathogen has not spread to natural forests. The disease was apparently eradicated in that nursery, preventing propagation of the hybrid. In Oregon, three lineages (NA1, NA2 and EU1) of *P. ramorum* have now spread to natural forests. This generates the potential for hybridization among those lineages in forests; the spread of a hybrid lineage would be more difficult to contain in natural forests than in nurseries where containment and eradication are possible. Our study highlights the importance of genomic biosurveillance for the detection of plant pathogen variants and hybrids to inform mitigation strategies.

**Methods**

**Sample collection and genomic characterization.** We obtained isolates of *P. ramorum* from nurseries in British Columbia (BC), Canada between 1995 and 2017 (Supplementary Table 2 and Supplementary Data 1). Leaf samples showing lesions were tested by qPCR for presence of *P. ramorum* genomes. Data is available from NCBI bioprojects PRJNA791184, PRJNA427329, PRJNA559872, PRJNA177509, and PRJNA558041 (https://www.ncbi.nlm.nih.gov/bioproject). More information is provided in Supplementary Data 1.

**Variant calling and filtering.** Filtered reads from the genomes generated in this study and those obtained from other sources were mapped to the *P. ramorum* NA1 lineage reference genome PR-102_v3.1 using bwa mem version 0.7.1753 with
default settings. Mapped reads were sorted using SAMtools version 1.9^{43} and genome-wide average coverage was assessed using Qualimap version 2.2.1^{43} with the bamqc option. Duplicated reads were annotated using the s (https://gatk. broadinstitute.org/hc/en-us). Variant calling was performed in two steps using Genome Analysis Toolkit (GATKVersion 4.1.0.0). First, variants were called for each individual genome using GATK’s HaplotypeCaller with the -ERC GVCF option and the ploidy was set to 2. Next, variant files were merged using GATK’s CombineGVCFs function and then joint variant calling across all individuals was performed with GenotypeGVCFs. Raw variants were reduced to SNPs only and filtered using GATK’s SelectVariants and VariantFiltration functionalities. Only SNPs satisfying the following criteria were retained: QD ≥ 500, QD ≥ 10, MQ ≥ 40, MQ ≤ 10 and MQRankSum, RankQD, RankSVM, ReadPosRankSum ≥ −2 and ≤2. Read depth (DP) was restricted to ≥1000 (≥100 percentile) and ≤10000 (≥99 percentile) to further increase SNP quality and only biallelic SNPs were retained. Next, SNPs were filtered on the genotype level using VCFTools version 0.1.16 (https://vcftools.github.io/man_latest.html). Genotypes with a read depth <10 (−minDP 10) and a genotype quality <30 (−minQ 30) were set to no-call, and variants with more than 5% missing genotypes were removed. Finally, invariant sites across all samples were removed. For some analyses (e.g., minimum spanning network, admixture) the SNP set was further restricted to variants without missing genotypes.

Extraction of mitochondrial haplotypes. Trimmed reads were mapped to the mitochondrial contig of the NA1 lineage reference genome PR-102_v3.1^{18} (NCBI accession: NC_009384.157) and variants called as described for the nuclear genome. Variants were restricted to SNPs only and the distribution of “variant confidence normalized by unfiltered depth of variant samples (QD)” was assessed to confirm the homogenous state across individuals. SNPs were filtered using GATK’s SelectVariants and VariantFiltration functionalities with the following criteria: QD ≥ 20, QD ≤ 10 and only bi-allelic SNPs with no missing genotypes were accepted. VCFTools was then used to set genotypes with low support (read depth <10, i.e. −minDP 10 and genotype quality <40, i.e. −minQ 40) to no-call. Sites with more than 5% missing genotypes and invariant SNPs across all individuals were removed.

Population genomics analyses. Population structure and genetic connectivity of the unlinked isolates with the four lineages of *P. ramorum* was evaluated using principal component analysis (PCA) with SNPrelate (version 1.28)^{38}. Linkage disequilibrium-pruning was applied on the VCF dataset of 450,656 SNP loci to subsample 3,780 markers with reduced linkage (R^2 < 0.20). Ancestry estimation was conducted using Admixture among the *P. ramorum* lineages with the full SNP set with K = 4, the number of lineages found in North America and Europe. To assess reticulated relationships between *P. ramorum* individuals a neighbor-net phylogenetic network was reconstructed using SplitsTree version 4.12.8^{39} with pairwise Nei’s genetic distances with the R package STAMP (version 1.6.3)^{40}.

Genome-wide hybridization detection and gene flow simulations. We used the function hybridization in the R package Alegenet (version 2.1.5) to simulate 10 F1 hybrids between NA2 and EU1 and 10 backcrosses of the hybrid to the EU1 or NA2 parent. We generated a pairwise distance matrix with the dist function hyDe testing was performed with the *P. ramorum* reference genome (PR-102, v3.1) and the giff file to build the annotation database. Functional class consisted of missense (non-synonymous mutations); nonsense mutations that cause stop codon loss or gain and frameshifts and silent mutations are synonymous. High impact mutations result in stop codon loss and gain and frameshifts; moderate impact mutations cause a codon to change the amino acid; low modifier mutations are in codons that produce the same amino acid; modifier mutations are in introns, upstream or downstream regions.

We measured growth of isolates of *P. ramorum* lineages and the EU1 × NA2 hybrids on carrot agar (CA) plates for four days at 20 °C; 7-mm plugs were transferred to fresh CA plates (3 replicates) and incubated in the dark at 20 °C. Isolates of the putative parental lineage NA2 (n = 11), EU1 (n = 11), and two hybrid isolates were used in the growth experiments. Host inoculations were performed in triplicate, using different rhododendron leaves. Rhododendron leaves were inoculated by placing a 7-mm plug from a 7-day-old colony of *P. ramorum* in the center of a detached leaf and measuring lesion development after 3, 4, 5, 6 and 10 days. The original data is available in Supplementary Data 2.

Statistics and reproducibility. For the HyDe analysis, we obtained average y-values with standard deviations by subsampling the full dataset of 450,656 SNP loci 500 times for 10,000 random loci. The Z-statistic was used to test for significance of the y-values, and x-values were adjusted for multiple testing using Bonferroni adjustment (α = 0.05 x 20). The Z-statistic was calculated using the -z functionality of R and the results were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate correction. The null hypothesis that there is no difference in effect size was tested with a two-tailed paired t-test (α = 0.05). The growth and inoculation experiments (shown in Fig. 2) were done in triplicate using 11 different isolates of the pathogen for the NA2 and EU1 lineages and two hybrid isolates. The colony diameters were measured after 2, 3, 6, 7, and 9 days and mean daily radial growth rate was calculated using linear regression (log[log(radius)] – hours) in R. For the host necrosis area, the necrosis area was measured in mm^2 and averaged. One-way ANOVA with multiple comparisons using Tukey contrasts for multiple comparison of means were performed between lineages for growth and lesion size.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The raw number of reads supports the findings of this study are included in the paper and all sequence data are deposited in NCBI GenBank as bioprojects, PRJNA791184, PRJNA427339, PRJNAS569872, PRJNA177509, and PRJNA558841. Source data underlying Fig. 3 are presented in Supplementary Data 2.

Code availability. The code used in this study is available on GitHub (https://github.com/ Feaunicoh/SOD_hybrid_tree/v1.0) and Zenodo with https://doi.org/10.5281/zenodo.6460530.
58. Zheng, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328 (2012).

59. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267 (2006).

60. Pemberton, L. W., Cogan, N. O. & Forster, J. W. STAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol. Ecol. Resour.* 13, 946–952 (2013).

61. Blischak, P. D., Chifman, J., Wolfe, A. D. & Kubatko, L. S. HyDe: a Python package for genome-scale hybridization detection. *Syst. Biol.* 67, 821–829 (2018).

62. Patterson, M. et al. WhatsHap: weighted haplotype assembly for future-generation sequencing reads. *J. Comput. Biol.* 22, 498–509 (2015).

63. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313 (2014).

64. Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830 (2011).

65. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92 (2012).

Acknowledgements
This work was supported by Genome Canada, Genome British Columbia, Genome Quebec, Natural Resources Canada, the Canadian Food Inspection Agency and FPInnovations through the Large Scale Applied Research Project (LSARP 10106. The authors acknowledge the help of Debbie Shearlaw, Miranda Newton, and Marie-Claude Gagnon from CFIA for technical assistance and Lucyna Kumor from Plant and Seed Pathology Laboratory at CFIA for the cultures.

Author contributions
R.C.H., G.J.B., N.F., A.C., and R.H. conceived the project, obtained the samples, and performed the genome analyses. K.H., E.D., A.L.D., A.C., S.K., and E.G. performed the phenotyping. N.C.C. and N.J.G. generated a reference genome that was used for mapping and analyses and provided annotations. R.C.H., N.F., R.H., and G.J.B. wrote the manuscript and the supplementary information with input from all other authors.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-022-03394-w.

Correspondence and requests for materials should be addressed to Richard C. Hamelin or Nicolas Feau.

Peer review information Communications Biology thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Diane Saunders, Caitlin Karniski and George Inglis. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022