Package 'pinfsc50'

July 23, 2025

Title Sequence ('FASTA'), Annotation ('GFF') and Variants ('VCF') for 17 Samples of 'P. Infestans" and 1 'P. Mirabilis'

Version 1.3.0

Maintainer Brian J. Knaus <briank.lists@gmail.com>

Description Genomic data for the plant pathogen ``Phytophthora infestans.'' It includes a variant file ('VCF'), a sequence file ('FASTA') and an annotation file ('GFF'). This package is intended to be used as example data for packages that work with genomic data.

Depends R (>= 3.2.1)

License GPL

RoxygenNote 7.2.3

Encoding UTF-8

NeedsCompilation no

Author Brian J. Knaus [cre, aut],
     Niklaus J. Grunwald [aut]

Repository CRAN

Contents

---

pinfsc50_data	*pinfsc50: A package containing the sequence, annotations and variants for* P. infestans *Supercontig_1.50.*

Description

The pinfsc50 package contains data from *Phytophthora infestans* intended to be used as example data.

pinfsc50 functions

This package contains no functions.

Files in inst/extdata

pinf_sc50.fasta - FASTA format file containing the nucleotide sequence for *Phytophthora infestans* T30-4 Supercontig_1.50. This data was published in Haas et al. (2009).

pinf_sc50.gff - GFF format file containing annotations for Supercontig_1.50

pinf_sc50.vcf.gz - gzipped VCF format file containing variant information for Supercontig_1.50. This file was created with the GATK's haplotype caller. The data were then phased with beagle4. beagle4 returns a VCF file which lacks much of the diagnostic information contained in the input file. I therefore stripped the unphased genotypes from the original file and pasted on the phased genotypes from beagle4 to create a vcf file with the genotypes from beagle4, but all of the other information contained in the GATK's haplotype caller's file. The goal here was to create a dataset that could provide a diverse set of examples for learning how to work with this type of data. I would not consider this a 'best practices' for an actual data analysis workflow.

Short read data for sample t30-4 was downloaded from the Broad's *P. infestans* page. Short read data from sample blue13 was published in Cooke et al. (2012). Short read data from samples DDR7602, LBUS5, NL07434, P10127, P10650, P11633, P12204, P13527, P1362, P13626, P17777us22, P6096 and P7722 were published in Yoshida et al. (2013). Short read data from samples BL2009P4_us23, IN2009T1_us22, RS2009P1_us8 were published in Martin et al (2013).

References

Cooke, D. E., Cano, L. M., Raffaele, S., Bain, R. A., Cooke, L. R., Etherington, G. J., ... & Kamoun, S. (2012). Genome analyses of an aggressive and invasive lineage of the Irish potato famine pathogen. PLoS Pathog, 8(10), e1002940.

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H., Handsaker, R. E., Cano, L. M., ... & Liu, Z. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature, 461(7262), 393-398.

Martin, M. D., Cappellini, E., Samaniego, J. A., Zepeda, M. L., Campos, P. F., Seguin-Orlando, A., ... & Gilbert, M. T. P. (2013). Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. Nature communications, 4.

Phytophthora infestans Sequencing Project, Broad Institute of Harvard and MIT (https://www.broadinstitute.org/).

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., ... & Burbano, H. A. (2013). The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. Elife, 2, e00731.

Examples

```
## Not run:
dna <- system.file("extdata", "pinf_sc50.fasta", package = "pinfsc50")
dna <- ape::read.dna(dna, format="fasta")
gff <- system.file("extdata", "pinf_sc50.gff", package = "pinfsc50")
gff <- read.table(gff, header=FALSE, sep="\t", quote = "")
vcf <- system.file("extdata", "pinf_sc50.vcf.gz", package = "pinfsc50")
```

```
vcf <- vcfR::read.vcfR(vcf)

## End(Not run)
```

Index