Mitochondrial genome sequences and comparative genomics of *Phytophthora ramorum* and *P. sojae*

Martin, F.N.[1], Morgan, J.[2], Putnam, N.[2], Tyler, B.[4], Boore, J.L.[2,3]

[1]USDA-ARS, Salinas, CA, [2]DOE Joint Genome Institute and Lawrence Berkeley National Laboratory, Walnut Creek, CA, [3]University of California, Berkeley, CA, [4]Virginia Bioinformatics Institute, Blacksburg, VA

Corresponding author: Frank Martin
USDA-ARS
Salinas, CA 93905
fmartin@pw.rs.usda.gov

ABSTRACT

The complete sequences of the mitochondrial genomes of the oomycetes of *Phytophthora ramorum* and *P. sojae* were determined during the course of their complete nuclear genome sequencing (Tyler et al. 2006). Both are circular, with sizes of 39,314 bp for *P. ramorum* and 42,975 bp for *P. sojae*. Each contains a total of 37 identifiable protein-encoding genes, 25 or 26 tRNAs (*P. sojae* and *P. ramorum*, respectively) specifying 19 amino acids, and a variable number of ORFs (7 for *P. ramorum* and 12 for *P. sojae*) which are potentially additional functional genes. Non-coding regions comprise approximately 11.5% and 18.4% of the genomes of *P. ramorum* and *P. sojae*, respectively. Relative to *P. sojae*, there is an inverted repeat of 1,150 bp in *P. ramorum* that includes an unassigned unique ORF, a tRNA gene, and adjacent non-coding sequences, but otherwise the gene order in both species is identical. Comparisons of these genomes with published sequences of the *P. infestans* mitochondrial genome reveals a number of similarities, but the gene order in *P. infestans* differs in two adjacent locations due to inversions. Sequence alignments of the three genomes indicated sequence conservation ranging from 75 to 85% and that specific regions were more variable than others.

Keywords: inverted repeat, *Phytophthora infestans*

INTRODUCTION

The genus *Phytophthora* has a wide geographic distribution throughout the world and contains more than 70 species, many of which cause important plant diseases (Erwin and Ribiero 1996). While members of the genus, along with other oomycetes, share morphological similarities with eumycotian fungi, these features have arisen independently, because oomycetes are more closely related phylogenetically to chromophyte algae within the larger group Stramenopiles (Förster et al. 1990a; Knoll 1992; Baldauf and Palmer 1993; Wainright et al. 1993; Bhattacharya and Stickel 1994; Weerakoon et al. 1998; Dick 2001). Members of this genus differ from eumycotan fungi in features such as being diploid throughout their life cycle and formation of motile, biflagellate spores called zoospores that are capable of swimming in water.

The mitochondrial genomes in the genus have been reported to be circular and to range in size from approximately 37.0 to 45.3 kb (Paquin et al. 1997, Avila-Adame et al. 2005; McNabb and Klassen 1988, Förster et al. 1987, Shumard-Hudspeth and Hudspeth 1990). These have been commonly used in RFLP studies for identification of isolates and to help clarify the taxonomic placement of particular species (reviewed in Erwin and Ribeiro 1996). Mitochondrial gene sequences also have been used to infer phylogenetic relationships among species of the genus as well (Martin and Tooley 2003, 2004; Kroon et al. 2004). The only species of *Phytophthora* for which we have a complete mitochondrial genome sequence is *P. infestans*, the causal agent of potato late blight. This has been determined for four separate haplotypes, which range in size from 37,957 bp for the Ia haplotype (Paquin et al. 1997) and from 37,992 to 39,870 bp for the Ia, IIa,

and IIb haplotypes (Avila-Adame et al. 2005). A total of 67 shared coding regions were identified in these genomes encoding for mitochondrial respiratory chain proteins, subunits of the mitoribosome, ribosomal RNAs, tRNAs, and unassigned ORFs. This same set of coding regions were also found in the related Peronosporomycete *Saprolegnia ferax* (Grayburn et al. 2004). Comparisons among the different *P. infestans* haplotypes indicate that intraspecific variation is due to changes in nucleotide sequences dispersed throughout the genome as well as length mutations caused by insertions/deletions that occurred primarily in two locations. It is unclear how this intraspecific variation relates to interspecific variability and genome divergence among species.

Interest in the genus *Phytophthora* has increased lately due to the serious impact several species are having as plant pathogens. *Phytophthora ramorum* is a recently described species that initially was found to be responsible for diseases of nursery crops in Germany and the Netherlands (Werres et al. 2001). It has subsequently spread to other European countries and more recently has become a problem in field ecosystems (Brasier et al. 2005). While this species also is a problem in some nursery production crops in North America, a far bigger impact has been its role as the cause of sudden oak death, a disease that has killed large numbers of trees and shrubs in natural ecosystems in central coastal California (Rizzo et al. 2002, Davidson et al. 2003). This is a highly regulated pathogen with stringent quarantine restrictions in place in North America and Europe in an effort to halt its spread. *Phytophthora sojae* is widely spread in soybean (*Glycine max*) production areas of North America and Australia and causes serious crop

production losses due to root and stem rot (Erwin and Ribiero 1996). There is a continuing effort in soybean breeding programs to develop resistant germplasm as a means for controlling the disease.

Complete draft sequences for the nuclear genomes of *P. ramorum* and *P. sojae* have been recently determined (Tyler et al. 2006). As part of this sequencing project, the complete mitochondrial genomes were also assembled. The objective of this submission is to annotate and describe these mitochondrial genomes and compare them to the mitochondrial genomes of *P. infestans* and other Peronosporomycetes.

MATERIALS AND METHODS

Strains sequenced – Mitochondrial sequences were obtained from *P. ramorum* strain Pr-102 (isolated form California) and *P. sojae* strain P6497 (isolated from Mississippi). Sequences for the Ia haplotype mitochondrial genome of *P. infestans* (Paquin et al. 1997) were obtained form GenBank (NC002387), as were those recently available for haplogypes Ib, IIa, and IIb (AY894835, AY898627, AY898628; Avila-Adame et al. 2005) and the peronosporomycete *S. ferax* (AY534144; Grayburn et al. 2004)

Sequencing and contig assembly– For each of these two species, total DNA preparations were randomly sheared using a Hydroshear device (Gene Machines, location needed), in separate aliquots, to fragments averaging either about 3 kb or about 8 kb. These were gel purified and enzymatically repaired to blunt ends, then cloned into plasmids to generate two genomic libraries. An additional library was created in a

fosmid vector. End sequences were determined for a large number of randomly selected clones from each of these libraries, then assembled using JAZZ **(REF)** to form a complete draft whole-genome shotgun assembly **(REF)** of these nuclear genomes. Detailed protocols are available at **<WEBSITE>** and this process and the results will be further described in a manuscript reporting the complete nuclear genome sequences. Although no effort was expended to target the mitochondrial genomes, even a small contamination by mtDNA in these preparations, coupled with the high molarity of these sequences compared to any portion of the nuclear genomes, guarantees that any whole-genome shotgun sequencing projects will include many sequencing reads from clones of mtDNA. **ADD STATISTICS SPECIFIC FOR THE MTDNAS ON THE NUMBER OF READS FROM EACH LIBRARY, THE DEPTH OF COVERAGE, AND THE OVERALL QUALITY OF THE ASSEMBLY AND OF THE CONSENSUS SEQUENCE.**

Annotation and comparative genomics - Annotation of coding regions was done using DS Gene v1.5 (Accelrys, San Diego, CA). Identification of protein- and rRNA-encoding genes was done by comparison with sequences reported for *P. infestans* (Paquin et al. 1997; NC002387) and BLAST analysis to other sequences in GenBank. Genes for tRNAs were found using tRNAscan SE v1.1 (Lowe and Eddy 1997; http://www.genetics.wustl.edu/eddy/tRNAscan-SE/). Comparisons among genomes was done using mVISTA (Mayor et al. 2000, Frazer et al. 2004; http://genome.lbl.gov/vista/servers.shtml). Sequence alignments within mVISTA were done using LAGAN (Brudno et al. 2003).

RESULTS AND DISCUSSION

Genome size and organization

The mitochondrial genomes for both species are circular and range in size from 39,314 bp for *P. ramorum* (Fig. 1; GenBank XXX) to 42,975 bp for *P. sojae* (Fig. 2; GenBank XXX) with a %GC content of 22.0% and 21.7%, respectively. This compares to 37,957 bp for the Ia haplotype (Paquin et al 1997) and 37,992, 39,870 and 39,840 bp for the Ia, IIa, and IIb haplotypes of *P. infestans*, respectively (Avila-Adame et al. 2005). The 37 protein-and rRNA-encoding genes found in *P. infestans* mtDNA (Paquin et al. 1997) are also present in each of these two other species and similar to *P. infestans*, none contain introns. This set comprises 18 respiratory chain proteins, 16 ribosomal proteins, the rRNAs for the large and small ribosomal subunits, and an import protein *(secY)* (Fig. 3). ATG was the start codon for all genes and with the exception of *nad11* (TGA), the termination codon for all assigned genes is TAA. This gene had the same termination codon in *P. infestans* (Paquin et al. 1997), but in *S. ferax* it was TAA (Grayburn et al. 2005).

There are a total of 19 amino acids encoded with the same 25 tRNAs as reported for *P. infestans* (Paquin et al. 1997) including multiple tRNAs for *trnG* (GCC, UCC), *trnL* (UAA, UAG), *trnR* (UCU, GCG), *trnS* (GCU, UGA) and three copies of *trnM* (CAU) identified by tRNA Scan. But one of these $trnM_{CAU}$ is identical to what was classified as $trnI_{CAU}$ in *P. infestans* and *S. ferax* (Paquin et al. 1997, Grayburn et al. 2004). For both these species it was determined that in view of the conclusions of Gray et al. (1998) that

this tRNA is actually a $trnI_{CAU}$ due to post transcriptional modification to lysidine to allow translation of the AUA codon for isoleucine. The other two copies of $trnM_{CAU}$ function in initiator and elongator roles. *TrnT* is not encoded in these genomes. *Phytophthora ramorum* has an additional copy of $trnR_{UCU}$ adjacent to *cob* relative to *P. sojae* due to this tRNA being encoded in the inverted repeat (

10.8 to 25% at an amino acid level. This compares to a sequence divergence for the *cox* 2 mitochondrially encoded gene of 5.9 to 6.0% at a nucleotide level and 1.6 to 3.5% at an amino acid level. It is questionable if *orf79* represents a functional gene since it has limited sequence conservation among *P. infestans, P. ramorum* and *P. sojae* at either a nucleotide or amino acid level (Table 2). There were six and one additional putative ORFs greater than 100 bp in *P. sojae* and *P. ramorum*, respectively, ranging in size from 294 to 621 bp that were not present in the other species. The six additional ORFs in *P. sojae* are present in two locations; *orf206* is between $trnY_{GUA}$ and the *rrnS* while the remaining five (*orf115, orf97,* $orf111_a$, $orf111_b$, and $orf100_b$) are clustered together between *nad6* and *nad4L* (Fig. 2). The 3' end of *orf115* overlaps the 5' end of *orf97* by 20 bp and the 3' end of *orf97* overlaps $orf111_a$ by 17 bp. There is only one unique ORF (*orf175*) longer that 100 bp in *P. ramorum* mtDNA, and this is part of the inverted repeat and is present in two copies in opposite orientation (Fig. 1 and discussed below). The termination codon for all of these unassigned ORFs is TAA with the exception of *orf97* and $orf111_b$ (TAG) and $orf111_a$ (TGA). Both these termination codons are used in different ORFs in haplotype IIa and IIb of *P. infestans* (Avila-Adame et al. 2005). BLAST analysis of sequences in GenBank did not identify any sequence homology among these ORFs or potential homologs for these putative coding regions.

The coding regions are closely packed in the genome with 70% of the spacer regions less than 30 bp long (Table 3). Overall the spacer regions represent a relatively small percentage of the genome compared to coding regions (11.5% and 18.4% for *P. ramorum* and *P. sojae*, respectively). Genes are divided between the two strands and in general are

clustered into five non-overlapping groups alternating between strands (Fig. 1 and 2). Starting at base 1 the first group is *rrnL* through *cox1* followed by *trnR* through *nad 2*, *nad7* through *nad6*, *nad4L* through *trnL*, and *rpl2* to *orf100*. There are some exceptions to these groupings, most notably transcription of *orf206* in *P. sojae* and *secY* through *rps13* for e c i s c q 0.24 0 0 0.24d

Inverted repeat

One unique feature of the *P. ramorum* genome is that it contains an inverted repeat (IR) of 1,150 bp located in one case between *cox1* and *cob* (bases 9,540 to 10,689) and again in opposite orientation between *nad6* and *nad4L* (bases 26,173 to 27,322). This inverted repeat contains *orf175*, which is unique to this species, as well as $trnR_{UCU}$. The first copy of the IR starts eight bp after the termination codon of *cox1* and the opposite end includes 13 bp of the 3' end of *cob*. The second copy starts with 74 bp of the 3' end of the *nad6* gene and has 38 bp of the 3' end of *nad4L*. This copy of the IR is in the same position of the genome as the clustered five unique ORFs of *P. sojae.* Comparisons of sequences between *cox1* and *cob* and *nad6* and *nad4L* for *P. ramorum* and *P. sojae* (the region where the IR is found in *P. ramorum*) revealed limited sequence similarity between these two species. Likewise, there was limited sequence similarity in comparison of region between *nad6* and *nad4L* from both these species and *P. infestans.* Without further analysis of a greater number of mitochondrial genomes in the genus it is unclear if the IR arose from duplication of a specific region of the genome or if its presence in a reduced state reflects a deletion of the large IR found in other genera in the Peronosporomycetes (discussed more below). However, given that the single copy of $trnR_{UCU}$ in *P. infestans* and *P. sojae* is present adjacent to the *nad6* gene, and one of the copies of this tRNA gene is in the same position for *P. ramorum*, it is possible that this location would reflect the ancestral position and the other arm of the IR between *cox1* and *cob* being the secondarily duplicated copy.

The only other example of a completely sequenced mitochondrial genome containing an IR is *S. ferax,* which has an IR of 8,618 bp, representing 37% of the genome size and encoding four proteins, five tRNA genes, and both rRNAs (Grayburn et al. 2004). These coding regions are transcribed from both strands of the mitochondrial genome, which is different from the two coding regions in the IR of *P. ramorum* transcribed in the same direction. One similarity between these genomes is that in *S. ferax*, one arm of both copies of the IR terminates with a partial sequence of a coding region (the 3' end of *nad5*), whereas in *P. ramorum*, three of the four termini end in a coding region. One end of the IR encodes the terminal 13 bp of the 3' end of *cob* and the same end of the second copy has 74 bp of the 3' end of *nad6*. The opposite end either terminates within the spacer region before the *cox1* gene in one copy or has 38 bp of the 3' end of *nad4L* in the other.

The presence of an IR in *P. ramorum* is unusual for the genus *Phytophthora*. The only other example where this has been observed is in *P. megasperma* (Schumrd-Hudspeth and Hudspeth 1990). Based on restriction mapping and Southern analysis a short inverted repeat of 0.5 to 0.9 kb in size was identified with one copy adjacent to *cox2* and the other adjacent to the *cob/atp9* genes (this later position is similar to what was observed for one arm of the IR in *P. ramorum*). Inverted repeats in the mitochondrial genome are common in the closely related genus *Pythium* (McNabb et al. 1987, McNabb and Klassen 1988, Martin 1991, Martin 2000) and have also been found in other oomycetes such as *Achyla* spp. (Hudspeth et al. 1983, Boyd et al. 1984, Schumard et al. 1986), *Aplanopsis terrestris, Leptolegnia caudate* and *Sapromyces elongates* (McNabb

and Klassen 1988) and *Saprolegnia ferax* (Grayburn et al 2004). An IR also has been reported in the chytrid *Hypochytrium catenoides* (McNabb et al. 1988). In cases where an IR has been described this region represents reflect a larger proportion of the genome size (greater than 37% overall, but more than 71% for *Pythium* spp.) and contains the large and small ribosomal RNA coding regions, which was not the case for the IR observed with *P. ramorum*.

Genome comparisons

With the exception of the IR in *P. ramorum* the gene order in *P. sojae* was the same, however, with *P. infestans* there are two inversions relative to *P. ramorum* and *P. sojae* that have reversed the gene order in adjacent regions (Fig. 3). One inversion includes *cob, nad9* and *atp9* while the other is immediately adjacent and includes a total of 18 coding regions spanning from *nad3* to *atp1* (Fig. 3). The gene order in several regions also were found to be conserved with *S. ferax* (Fig. 3), including the linkage of *rps8, rpl6, rps2*, and *rps4* that Grayburn et al. (2004) noted was also conserved in the stramenopile *Chrysodidymus synuroides* (Chesnick et al. 2000). Although the gene order of more *Phytophthora* species needs to be examined to confirm this, the conservation of gene order in *P. ramorum* and *P. sojae* relative to *P. infestans* may be reflective of the evolutionary relationship among these species. In phylogenetic analysis using *cox2* and rDNA ITS data, *P. ramorum* and *P. sojae* were more closely grouped, with *P. infestans* being less so (Martin and Tooley 2004). A similar relationship was observed for analyses done with sequence data from the ITS region (Cooke et al. 2000) and data from ß-tubulin, elongation factor 1-α, *cox1*, and *nadh1* (Kroon et al. 2004).

Whole genome sequence alignments between *P. ramorum* and *P. sojae* revealed a sequence conservation of 83% and 76% identity, respectively (Table 4) with the lower value for *P. sojae* due to the larger genome size. Sequence alignments were also done with the type 1b mitochondrial haplotype of *P. infestans*, which is smaller than the other genomes at 37,957 bp (Paquin et al. 1997). The Ia haplotype of *P. infestans* has 85% sequence identity with the other two species whereas *P. ramorum* and *P. sojae* have 82% and 75.4% identity with *P. infestans*, respectively.

Genome comparisons using mVISTA provides a graphic representation of the variation among genomes. Using *P. ramorum* as the base sequence, comparison with *P. sojae* reveals a low level of sequence variation throughout the genome, but specific regions were more variable than others (Fig. 4). The most extensive variation was found in the spacer regions and in general, the larger the spacer region, the greater the sequence variability between species. For example, the spacer region between *orf79* and *cox2* was 446 bp long for *P. ramorum* (327 bp for *P. sojae*) and exhibited a low level of sequence conservation with *P. sojae* (Table 3, Fig. 4). Likewise, the regions between *trnY* and *rrnS*, *nad9* to *atp9*, *atp9* to *nad3*, and *atp1* to *nad5* have similar low levels of sequence conservation between species, as did the regions represented by the IR in *P. ramorum*. However, this association between spacer length and sequence variation was not always observed. The spacer region between *nad7* and *orf142* is 42 bp for *P. ramorum* (303 bp for *P. sojae*) and has minimal sequence conservation while the spacer between *trnL* and *nad11* is 144 bp for *P. ramorum* (149 bp for *P. sojae*) and has a sequence conservation of

79%. Likewise, the spacer between *nad2* and *nad7* is 116 bp for *P. ramorum* (113 bp for *P. sojae*) and has a sequence conservation of 84.4%. Vista comparisons including the mitochondrial genome of *P. infestans* (with specific regions reverse complemented to account for the inversions) gave results that did not differ appreciably from those observed in Fig. 4 (data not shown). In intraspecific comparisons of the four mitochondrial haplotypes of *P. infestans*, the majority of the variation was observed between *trnY* and *rrnS* as well as downstream of *orf79* (the regions where insertions/deletions were observed; Avila-Adame et al. 2005), however, there is also a region of 25 bp in the spacer between *nad3* and *nad5* where there is only 32% sequence conservation between the IIb haplotype and the others (data not shown).

Some of the regions where high levels of sequence variability were observed in comparisons between *P. ramorum* and *P. sojae* corresponded to the location of differences in genomic organization among species. For example, the terminal regions of the genomic inversions observed with *P. infestans* corresponded to regions of low sequence similarity between *P. ramorum* and *P. sojae* (Fig. 4, between *cox1* - *cob*, *atp9* - *nad3*, and *atp1* - *nad5*). One of these regions, the area between *cox1* and *cob*, is where one arm of the inverted repeat is located in *P. ramorum*. The other arm is located between *nad6* and *nad4L*, which also was where the five unique ORFs in *P. sojae* are located. Furthermore, *orf206* in *P. sojae* is located between *trnY* and *rrnS*, which was the same location in the genome of *P. infestans* where length mutations associated with the major intraspecific differences in genomic sequences were observed (Avila-Adame et al. 2005). Another region of the *P. infestans* mitochondrial genome where smaller length

mutations were observed is downstream of *orf79* (34 and 36 bp in length), which is also a region of sequence variation in comparisons between *P. ramorum* and *P. sojae*. Interestingly, the spacer region in *P. infestans* between *nad3* and *nad5*, which is one juncture of the inversion relative to *P. ramorum* and *P. sojae*, is also variable in haplotype IIb relative to the other three haplotypes. From comparing the gene maps for these three species it is interesting to note that two of the regions variable in interspecific comparisons (between *cox1* and *cob* and *nad6* and *nad4L*) correspond to the head-to-head juncture of two clusters of genes transcribed from opposite directions.

The results thus far suggest a high degree on gene order conservation in the genus *Phytophthora* with the differences observed explained by two inversions. One reason for this may be the large percentage of the genome represented by coding regions and the small sizes of the intervening spacer regions conferring some level of genome stability. When interspecific variation is observed (rearrangements, unique ORFs, IR) these tend to be found in specific locations in the genome where there are low levels of interspecific sequence conservation in the spacer regions. However, before firm conclusions about genome stability can be drawn additional comparisons among more species is needed to clarify this. This would also clarify the relationship between changes in genome organization and phylogeny in the genus.

ACKNOWLEDGEMENTS

This work was support by National Science Foundation Grant MCB-0242131 and U.S. Department of Agriculture Grant XXXXXXXXXXXXX and was performed partly under

the auspices of the U. S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract N

LITERATURE CITED

Avila-Adame C, Gomez-Alpizar L, Zismann V, Jones KM, Buell CR, Beagle Ristaino J (2006) Mitochondrial genome sequences and molecular evolution of the Irish potato famine pahtogen, *Phytophthora infestans*. Curr Genet 49:39-46.

Baldauf SL, Palmer JD (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. Proc Natl Acad Sci U S A 90:11558-11562.

Bhattacharya D, Stickel SK (1994) Sequence analysis of duplicated actin genes in *Lagenidium giganteum* and *Pythium irregulare* (Oomycota). J Molec Evol 39:56-61.

Boyd DA, Hobman TC, Gruenke SA, Klassen GR (1984) Evolutionary stability of mitochondrial DNA organization in *Achlya*. Can J Biochem Cell Biol 62:571-576.

Brasier CM, Denman S, Brown A, Webber J (2005) Sudden oak death (*Phytophthora ramorum*) discovered in trees in Europe. Mycol Res 108:1108-1110.

Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S (2003) NISC Comparative Sequencing Program. LAGAN and Multi-LAGAN: Efficient Tools for Large-Scale Multiple Alignment of Genomic DNA. Genome Res 13:721-731.

Cooke DEL, Drenth A, Duncan JM, Wagels G, Brasier CM (2000) A molecular phylogeny of *Phytophthora* and related oomycetes. Fungal Genet Biol 30:17-32.

Davidson JM, Werres S, Garbelotto M, Hanson E, Rizzo DM (2003) Sudden Oak Death and associated diseases caused by *Phytophthora ramorum*. Plant Health Progress doi:10.1094/PHP-2003-0707-01-DG.

1 Dick MW (2001) Straminipilous Fungi. Kluwer Academic Publishers.

2 Erwin DC, Ribeiro OK (1996) *Phytophthora* Diseases Worldwide. American
3 Phytopathological Society Press St. Paul, MN. 562 pp.

4 Förster H, Coffey MD, Elwood H, Sogin ML (1990) Sequence analysis of the small
5 subunit ribosomal RNAs of three zoosporic fungi and implications for fungal evolution.
6 Mycologia 82:306-312.

7 Förster H, Kinscherf, TG, Leong, S, and Maxwell, DP (1987) Molecular analysis of the
8 mitochondrial genome of *Phytophthora*. Curr Genet 12:215-218.

9 Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational
10 tools for comparative genomics. Nuc Acids Res July 1:32:W273-Ω279.

11 Grayburn WS, Hudspeth DSS, Gane MK, Hudspeth MES (2004) The mitochondrial
12 genome of *Saprolegnia ferax*: organization, gene content, and nucleotide sequence.
13 Mycologia 96:980-987.

14 Hudspeth MES, Shumard DS, Bradford JR, Grossman LI (1983) Organization of *Achlya*
15 mtDNA: A population with two orientation and a large inverted repeat containing the
16 rRNA genes. Proc Natl Acad Sci U S A 80:142-146.

17 Knoll HA (1992) The early evolution of eukaryotes: a geological perspective. Science
18 256:622-627.

1 Kroon LPNM, Bakker FT, van den Bosch GB, Bonnants PJ, Flier WG (2004)
2 Phylogenetic analysis of *Phytophthora* species based on mitochondrial and nuclear DNA
3 sequences. Fungal Genet Biol 41:766-782.

4 Martin FN (1991) Linear mitochondrial molecules and intraspecific mitochondrial
5 genome stability in a species of *Pythium*. Genome 34:156-162.

Martin FN (2000) Phylogenetic re

McNabb SA, Klassen GR (1988) Uniformity of mitochondrial DNA complexity in Oomycetes and the evolution of the inverted repeat. Exp Mycol 12:233-242.

McNabb SA, Eros RW, Klassen GR (1988) Presence and absence of large inverted repeats in the mitochondrial DNA of Hyphochytriomycetes. Can J Bot 66:2377-2379.

Paquin B, Laforest M-J, Forget L, Roewer I, Wang Z, Longcore J, Lang BF (1997) The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. Curr Genet 31:380-395.

Rizzo DM, Garbelotto M, Davidson JM, Slaughter GW, Koike ST (2002) *Phytophthora ramorum* as the cause of extensive mortality of *Quercus* spp. and *Lithocarpus densiflorus* in California. Plant Dis 86:205-214.

Shumard-Hudspeth DS, Hudspeth MES (1990) Genetic rearrangements in *Phytophthora* mitochondrial DNA. Curr Genet 17:413-415.

Shumard DS, Grossman LI, Hudspeth MES (1986) *Achlya* mitochondrial DNA: gene localization and analysis of inverted repeats. Mol Gen Genet 202:16-23.

Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang R, Aerts A, Damasceno CMB, Dou D, Dubchak I, Gijzen M, Gordon S, Govers F, Grunwald N, Huang W, Ivors K, Kamoun S, Krampis K, Lamour K, McDonald WH, Medina M, Meijer H, Nordberg E, Ospina-Giraldo MD, Morris P, Putnam N, Rash S, Rose JKC, Sakihama Y, Salamov A, Savidor A, Smith B, Smith J, Sobral BWS, Terry A, Torto-Alalibo T, Win J, Zhang H, Grigoriev I, Rokhsar D, and Boore J (2006) Genome Sequences of two *Phytophthora* species

responsible for Sudden Oak Death and Soybean Root Rot provide novel insights into their evolutionary origins and mechanisms of pathogenesis. Science: In review

Wainright PO, Hinkle G, Sogin ML, Stickel SK (1993) Monophyletic origins of the metazoa: an evolutionary link with fungi. Science 260:340-342.

Weerakoon ND, Roberts JK, Lehnen LP, Hardham AR (1998) Isolation and characterization of the single β-tublin gene in *Phytophthora cinnamomi*. Mycologia 90:85-95.

Werres S, Marwitz R, Man In'T Veld WA, Cock AWAM, Bonants PJM, Weerdt Md, Themann K, Ilieva E, Baayen RP (2001) *Phytophthora ramorum* sp. nov., a new pathogen on Rhododendron and Viburnum. Mycol Res 105:1155-1165.

Table 1. Differences in size (in bp) of specific genes among *Phytophthora infestans*, *P. ramorum*, and *P. sojae*.

Gene	*P. infestans*[a]	*P. ramorum*	*P. sojae*
cob	1,152	1,161	1,161
*rpl*5	534	528	534
*rps*3	804	816	831
*rps*7	474	432	432
*rps*11	417	420	417
*rps*13	414	417	414
*rps*19	234	237	237
*sec*Y	747	744	744

[a] Data from Paquin et al. 1997

Table 2. Sizes and % sequence divergence of open reading frames (ORF) shared among *Phytophthora infestans, P. ramorum, P. sojae*, and *Saprolegnia ferax*.

			P. ramorum		*P. sojae*		*S. ferax*[a]	
		bp	DNA	AA	DNA	AA	DNA	AA
Orf32	*P. infestans*[b]	99	15.2%	25%	11.1%	12.5%	-	-
	P. ramorum	99	-	-	8.1%	15.6%	-	-
	P. sojae	99	-	-	-	-	-	-
Orf64	*P. infestans*	204	7.8%	16.9%	5.9%	10.8%	30.9%	58.7%
	P. ramorum	198	-	-	6.1%	12.3%	28.8%	58.7%
	P. sojae	198	-	-	-	-	28.8%	58.7%
	S. ferax	195	-	-	-	-	-	-
Orf79	*P. infestans*	240	35%	53.2%	35.4%	61.04%	-	-
	P. ramorum	243	-	-	41.2%	55.1%	-	-
	P. sojae	243	-	-	-	-	-	-
Orf100	*P. infestans*	303	11.9%	18.0%	10.2%	16.0%	-	-
	P. ramorum	309	-	-	9.7%	17%	-	-
	P. sojae	303	-	-	-	-	-	-
Orf142	*P. infestans*	429	10.5%	20.1%	9.8%	16.2%	28.4%	53.24%
	P. ramorum	420	-	-	10.7%	18.7%	26.0%	49.6%
	P. sojae	429	-	-	-	-	27.3%	53.2%
	S. ferax	432	-	-	-	-	-	-
Orf217	*P. infestans*	654	11.8%	19.0%	8.9%	13.8%	-	-
	P. ramorum	660	-	-	9.7%	16.0%	-	-
	P. sojae	678	-	-	-	-	-	-
Cox 2	*P. infestans*	774	5.9%	1.6%	6.6%	2.7%	24.3%	26.2%
	P. ramorum	777	-	-	5.9%	3.5%	25.0%	27.0%
	P. sojae	777		-	-	-	25.1%	25.8%
	S. ferax	759	-	-	-	-	-	-

[a] Data from Grayburn et al. 2004

[b] Data from Paquin et al. 1997

Table 3. The sizes of spacer regions (bp) between coding regions and the percent sequence conservation in comparisons between *Phytophthora ramorum* and *P. sojae*.

Spacer region	Spacer length *P. ramorum*	Spacer region	Spacer length *P. sojae*	% Identity
rrnl to *trnN*$_{GUU}$	6		6	100%
trnN$_{GUU}$ to *trnS*$_{GCU}$	11		12	83.0%
trnS$_{GCU}$ to *trnM*$_{CAU}$	13		20	60.0%
trnM$_{CAU}$ to *trnP*$_{UGG}$	26		44	39.6%
trnP$_{UGG}$ to *trnM*$_{CAU}$	13		13	92.3%
trnM$_{CAU}$ to *rpl14*	18		20	80.0%
rpl14 to *rpl5*	6		6	100%
rpl5 to *trnG*$_{GCC}$	7		7	100%
trnG$_{GCC}$ to *trnG*$_{UCC}$	100		96	83.0%
trnG$_{UCC}$ to *trnY*$_{GUA}$	14		26	53.8%
trnY$_{GUA}$ to *rrns*	234	trnY$_{GUA}$ to orfB[a]	53	47.2%
		orfB to rns	159	65.0%
rrns to *trnW*$_{CCA}$	33		48	58.3%
trnW$_{CCA}$ to *orf79*	157		167	87.4%
orf79 to *cox2*	446		327	65.0%
cox2 to *orf32*	13		11	84.6%
orf32 to *cox1*	103		110	82.8%
cox1 to *cob*	1,145[b]		1,181	64.4%
cob to *nad9*	45		46	82.6%
nad9 to *atp9*	141		163	63.4%
atp9 to *nad3*	186		389	40.4%
nad3 to *trnD*$_{GUC}$	36		35	51.2%
trnD$_{GUC}$ to *atp6*	50		47	64.7%
atp6 to *cox3*	25		27	70.4
cox3 to *rps7*	69		60	79.7%
rps7 to *rps12*	-26		-26	overlap
rps12 to *trnV*$_{UAC}$	20		19	75.0%
trnV$_{UAC}$ to *trnI*$_{GAU}$	3		3	33.3%
trnI$_{GAU}$ to *trnQ*$_{UUG}$	1		1	100%
trnQ$_{UUG}$ to *trnR*$_{GCG}$	14		14	85.7%
trnR$_{GCG}$ to *rps10*	4		4	75.0%
rps10 to *trnF*$_{gaa}$	18		17	88.9%
trnF$_{GAA}$ to *nad2*	6		6	100%
nad2 to *nad7*	116		113	84.4%
nad7 to *orf142*	48		303	12.5%
orf142 to *trnH*$_{GUG}$	3		4	75.0%
trnH$_{GUG}$ to *nad4*	26		30	66.6%
nad4 to *trnE*$_{UUC}$	21		31	54.8%
trnE$_{UUC}$ to *atp1*	75		81	81.4%
atp1 to *nad5*	151		1,052	10.6%
nad5 to *nad6*	68		62	70.4%
nad6 to *trnR*$_{UCU}$	15		26	46.2%
trnR$_{UCU}$ to *nad4L*[c]	950		2,582	18.2%
nad4L to *nad1*	2		5	40.0%
nad1 to *nad11*	-4		-4	overlap
nad11 to *trnL*$_{UAG}$	143		149	79.0%
trnL$_{UAG}$ to *trnL*$_{UAA}$	9		9	66.6%
trnL$_{UAA}$ to *SecY*	17		20	60.0%
SecY to *orf64*	4		4	75.0%
orf64 to *trnC*$_{GCA}$	11		27	37.0

trnC$_{GCA}$ to *trnS*$_{UGA}$	4	5	80%
trnS$_{UGA}$ to *rps11*	10	10	100%
rps11 to *rps13*	12	12	91.7
rps13 to *rpl2*	28	28	100%
rpl2 to *rps19*	2	3	66.6%
rps19 to *rps3*	3	3	66.6%
rps3 to *rpl16*	2	2	100%
rpl16 to *trnM*$_{CAU}$	3	7	42.9%
trnM$_{CAU}$ to *orf217*	17	31	45.2%
orf217 to *atp8*	65	61	82.0%
atp8 to *trnK*$_{UUU}$	25	28	82.1%
trnK$_{UUU}$ to *trnA*$_{UGC}$	2	2	100%
trnA$_{UGC}$ to *rps14*	29	18	44.8%
rps14 to *rps8*	7	7	100%
rps8 to *rpl6*	5	8	50.0%
rpl6 to *rps2*	6	8	75.0%
rps2 to *rps4*	8	8	87.5%
rps4 to *orf100*	2	7	28.6%
orf100 to *rrnl*	280	27	9.5%

[a] In *P. sojae* there is a unique ORF (*orf*206) that splits this spacer region in two but it can still be aligned with the spacer region in *P. ramorum*. There is virtually no sequence homology between this spacer region in *P. ramorum* and *orf206* in *P. sojae*.

[b] This region in *P. ramorum* spans the inverted repeat and includes *orf175* and one copy of *trnR(ucu)* but is all spacer in *P. sojae*.

[c] In *P. ramorum* this region includes the second copy of the inverted repeat encoding *orf175* while in *P. sojae* it includes 5 ORFs (*orf115, orf97, orf111*$_a$*, orf111*$_b$*,* and *orf100*$_b$).

Table 4. Sequence conservation in whole mitochondrial genome comparisons with *Phytophthora infestans, P. ramorum* and *P. sojae*.

Genome comparisons[a]	CNS[b]	Sequence identity
P. ramorum vs *P. sojae*	89.0%	83.0% (*P. ramorum*)
		76.0% (*P. sojae*)
P. ramorum vs *P. infestans*	88.3%	82.0% (*P. ramorum*)
		84.4% *P. infestans*)
P. sojae vs *P. infestans*	91.6%	75.4% (*P. sojae*)
		85.0% (*P. infestans*)

[a] Mitochondrial genome sizes for *P. ramorum*, *P. sojae*, and *P. infestans* were 39,314 bp, 42,975 bp and 37,957 bp (NC002387, Paquin et al. 1997), respectively.

[b] CNS = sequence conservation above 75% over a 45 bp window in mVISTA

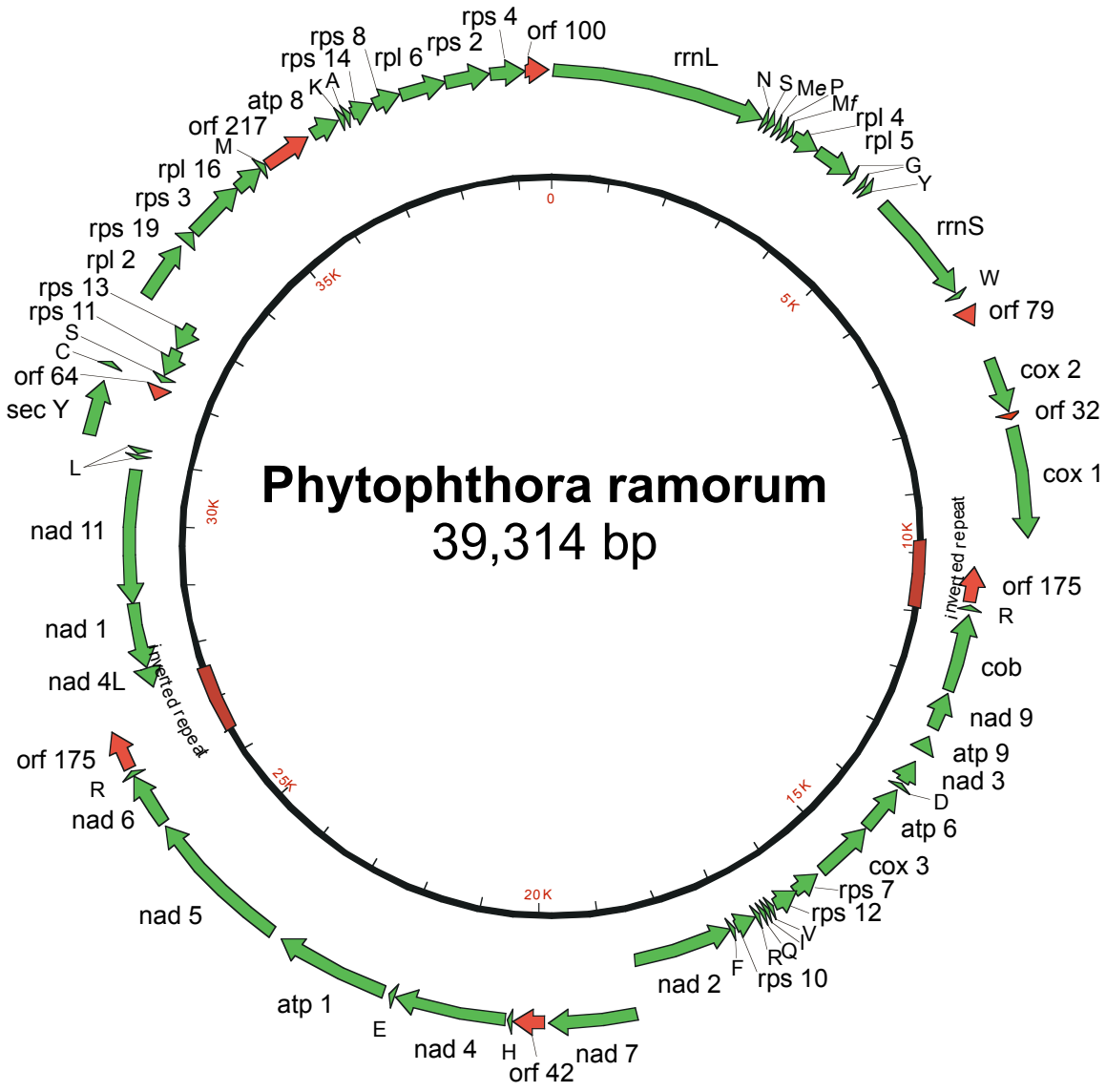
Figure legends

Figure 1. Mitochondrial gene map for *Phytophthora ramorum*. Arrows indicate transcriptional orientation, clockwise for the outer row and counter clockwise for the inner row with green representing coding regions and red putative ORFs. The position of the inverted repeat is indicated on the inner ring.

Figure 2. Mitochondrial gene map for *Phytophthora sojae*. Arrows indicate transcriptional orientation, clockwise for the outer row and counter clockwise for the inner row with green representing coding regions and red putative ORFs.

Figure 3. Conserved gene order in the mitochondrial genomes of *Phytophthora ramorum* and *P. sojae*. Comparisons with *P. infestans* and *Saprolegnia ferax* were based on Paquin et al. (1997) and Grayburn et al. (2004), respectively. The dashed line under *orf217* is to indicate that in *S. ferax* this region is represented by *orf273* (which has limited sequence conservation with *orf217*); otherwise the gene orde

regions that were inverted in the *P. infestans* mitochondrial genome relative to *P. ramorum* and *P. sojae.*

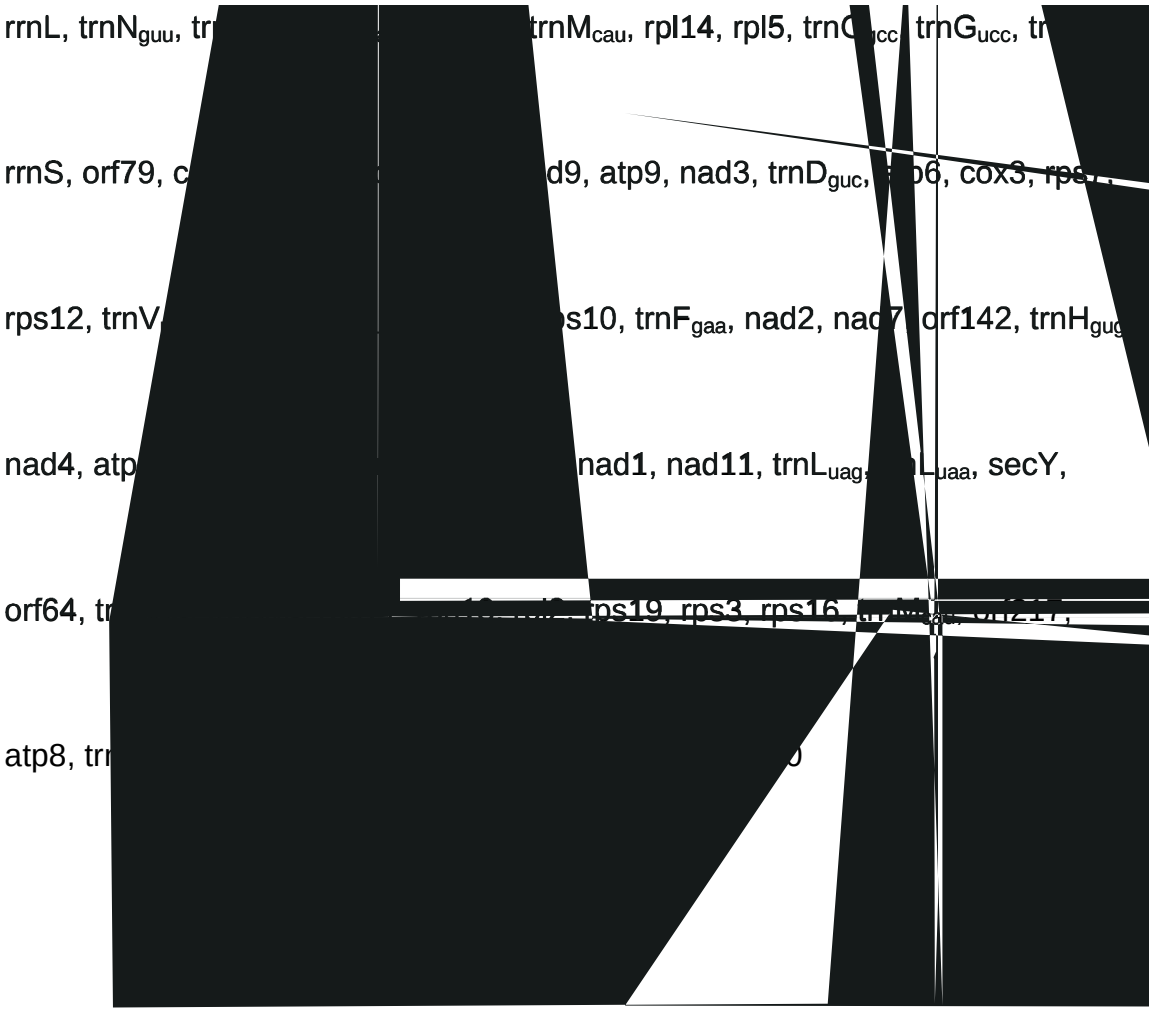

Phytophthora ramorum
39,314 bp
0
5K
10K
15K
20K
25K
30K
35K
rps 4
orf 100
rrnL
N
S
Me
P
Mf
rpl 4
rpl 5
G
Y
rrnS
W
orf 79
cox 2
orf 32
cox 1
orf 175
R
inverted repeat
cob
nad 9
atp 9
nad 3
D
atp 6
cox 3
rps 7
rps 12
V
Q
R
F
rps 10
nad 2
nad 7
orf 42
H
nad 4
E
atp 1
nad 5
nad 6
R
orf 175
inverted repeat
nad 4L
nad 1
nad 11
L
sec Y
orf 64
C
S
rps 11
rps 13
rpl 2
rps 19
rps 3
rpl 16
M
orf 217
atp 8
K
A
rps 14
rps 8
rpl 6
rps 2

1

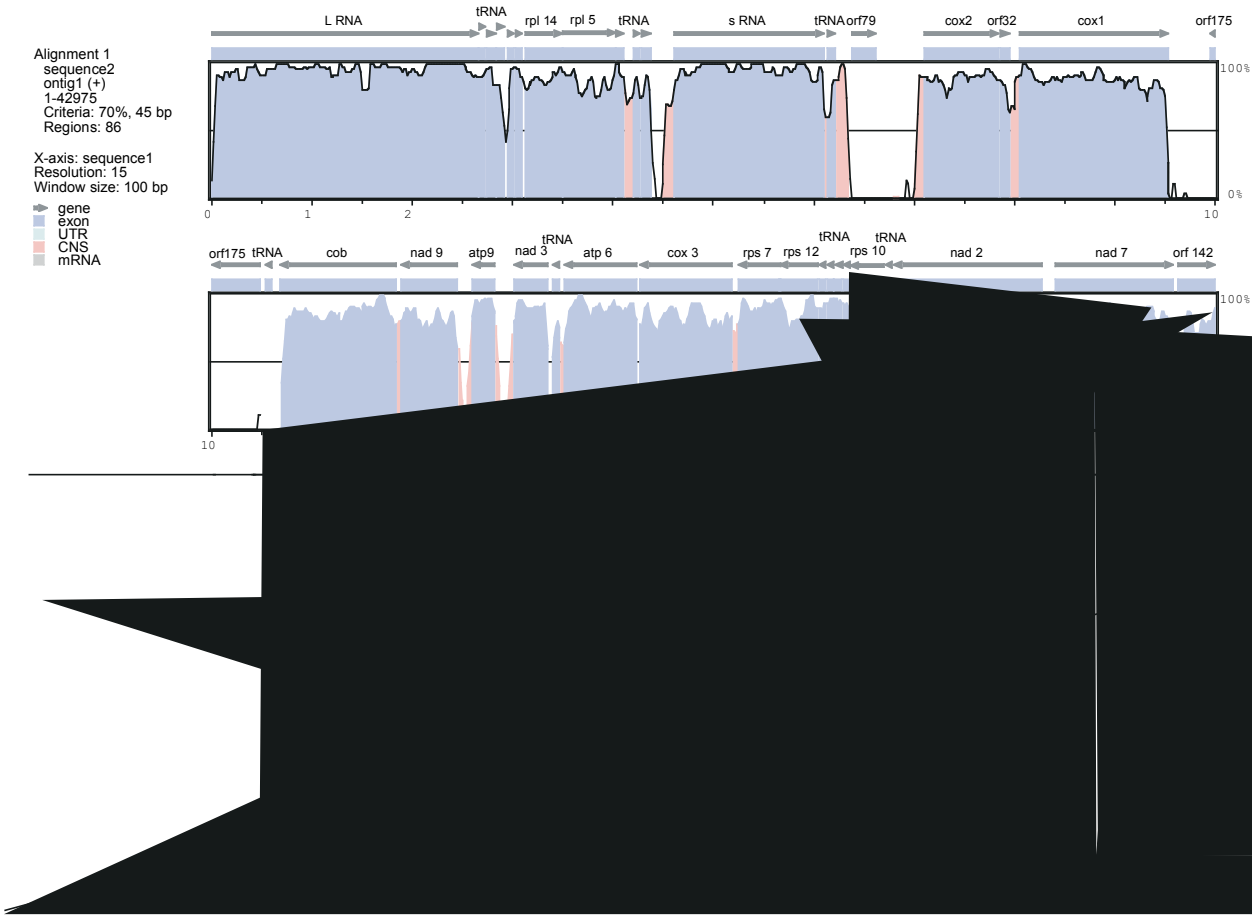
2

1
2

1


rrnL, trnN$_{guu}$, tr
trnM$_{cau}$, rpl14, rpl5, trnG$_{gcc}$, trnG$_{ucc}$, tr
rrnS, orf79, c
d9, atp9, nad3, trnD$_{guc}$,
6, cox3, rps
rps12, trnV
s10, trnF$_{gaa}$, nad2, nad7, orf142, trnH$_{gug}$
nad4, atp
nad1, nad11, trnL$_{uag}$,
L$_{uaa}$, secY,
orf64, tr
ps19, rps3, rps16,
atp8, tr

2
3


Alignment 1
sequence2
ontig1 (+)
1-42975
Criteria: 70%, 45 bp
Regions: 86
X-axis: sequence1
Resolution: 15
Window size: 100 bp
gene
exon
UTR
CNS
mRNA
L RNA
tRNA
rpl 14
rpl 5
tRNA
s RNA
tRNA
orf79
cox2
orf32
cox1
orf175
100%
0%
0
1
2
10
orf175
tRNA
cob
nad 9
atp9
nad 3
tRNA
atp 6
cox 3
rps 7
rps 12
tRNA
rps 10
tRNA
nad 2
nad 7
orf 142
100%
10