Intron Paper:

Outline:

Talk about homing endonucleases, the different types (LAGLIDADG, GIY-YIG, HNH etc.) move onto how they are present in both group I and group II introns in bacterial but also eukaryotic mitochondria and chloroplasts. Talk about these introns, they are self-splicing, encoding retrotransposable elements (i.e. endonucleases, maturases, retrotransposases), maybe a little about structure and prevalence in different genomes, humans and then say that group I have been found in lichens. Lichens are an awesome study system for symbiosis but also to look at the spread of these parasitic elements. We found widespread presence in these 58 genomes that have been assembled that encompass a large sampling of lichen mycobiont mitochondria. We report, number of genes and number of introns in each gene. We have clustered these intron sequences before aligning and see some general patterns that the introns in Usnea seem to all be related, etc. We focused in on the cox1 gene. We look at number of introns and number of these introns that contain homing endonucleases. We use a time calibrated tree to comment on the timescale that changes in these mitochondrial have undergone, such as genome size, number of genes, number of introns, number of homing endonucleases and suggest ancestral states for some of these variables. We note that the Usnea clade has extremely long branch lengths and is the most unusual with the overall highest number of introns but the fewest with homing endonucleases present (there is some literature showing that Usnea is highly divergent). This is in contrast to their conserved number of genes and perfect synteny across all the species examined.

**Introduction**

Group II introns are commonly found in both bacterial genomes and organellar genomes of lower eukaryotes but not in the more streamlined metazoan mitochondrial genomes (Lambowitz et al., 2004). The mitochondrion is believed to be derived from an endosymbiotic event of an early alpha-proteobacterium moving into a host cell and is one of the first examples of a symbiotic relationship (Gray et al., 1999). Group II introns have ribozyme activity, a common structure of six helical domains (Edgell et al., 2011) which helps to catalyze the splicing of exons directly next to them (Lambowitz et al., 2004) and often contain open reading frames (ORFs) with homology to proteins that have reverse transcriptase activity (Kennell et al., 1993; Michel & Lang 1985). The intron encoded protein can often be homologous to homing endonucleases (Matsuura et al., 1997). It is also believed that these mobile group II introns may be the presumed predecessor of spliceosomal introns and non-LTR retrotransposons in higher eukaryotes, which make up more than one third of the human genome (Lambowitz et al., 2004; Cordaux et al., 2009). It is estimated that in organellar genomes group II introns contain an ORF with homology to a retrotransposable element (maturase, endonuclease, or reverse transcriptase) up to one third of the time (Lambowitz et al., 2004; Rekadwad et al., 2017). Group II introns splice via two separate transesterification reactions that required an active site with Mg2+ present (Schmelzer et al., 1986; van der Veen et al., 1986).

These group II introns have homology to different retrotransposable elements: maturases that promote splicing (Wenzlau et al., 1989), reverse transcriptases that generate cDNA from RNA transcripts (Matsuura et al., 1997), and homing endonucleases which function as DNases (Lambowitz et al., 2004; Sethuraman et al., 2009; Hausner et al., 2012). Homing endonucleases are a rare-cutting class of DNases that are functionally and structurally distinct from restriction enzymes but behave in a similar manner (Sethuraman et al., 2009; Hausner et al., 2012). There are four main families of homing endonucleases that are identifiable using conserved amino acid sequences found in their functional domains (LAGLIDADG, GIY-YIG, His-Cys box, and HNH; Chevalier et al., 2001). These enzymes produce double stranded cuts at specific recognition sites, that can vary between 15-35 bp, in exons followed by double stranded repair and insertion of an intron containing allele (Chevalier and Stoddard 2001; Thiéry et al., 2010). Additionally, they are capable of horizontal transmission within populations and can impact overall genome size and structure (Belfort et al., 2014). Previous studies have shown that homing endonucleases play a key role in size polymorphisms between fungal mitochondrial genomes (Bhattacharya et al., 2002; Printzen & Ekman 2003; Haugen et al., 2004; Thiéry et al., 2010; Xavier et al., 2012; Mardanov et al., 2014; Kanzi et al., 2016).

Group II introns are often considered to be selfish parasitic elements that invade their host genomes. However, as described above the introns themselves can be parasitized by smaller ORFs with homology to proteins that promote intron mobility (Edgell et al., 2011). It has been suggested that these introns were originally ORF-less and that have been invaded multiple times over evolutionary history to make these composite parasitic elements (Toor et al., 2001; Edgell et al., 2011). A predominant theory is that these elements (group II intron and intron encoded protein) developed together rather than as two independent catalytic RNAs (Toor et al., 2001). Additionally, it is thought that introns that don’t contain ORFs are simply derivatives of their ORF containing cousins (Toor et al., 2001). Another, proposed alternative mechanism for intron invasion takes into mind that in order to avoid intron disruption the invading ORF would have to contain nearly the same nucleotides required to maintain the intricate intron folding (Edgell et al., 2011). Thus, the overlap of the ORF with an intron is thought to have come about by a process known as ‘core creep’ (Edgell et al., 2011). This is described as an extension of the ORF’s coding region by mutation of a stop codon into one specifying an amino acid. This results in the ORF being extended until the next in-frame stop codon (which is often connected with the gene that has been parasitized by the intron in which the ORF exists). This same strategy can be applied to core creep in the 5’ direction, producing overlap with the previous exon of the gene. The host genome often has repressive responses against these mobile group II introns because unregulation will lead to too high of a mutation load. Therefore, it has been thought for a while that these elements are simply selfish, parasitic elements. However, there have been recent studies that have shown that these introns can mobilize in response to stress-induced conditions (Coros et al., 2009; Robbins et al., 2011).

Previous work in Ascomycete fungi (e.g. Ophiostoma, Grosmannia) has demonstrated that the mitochondrial genome is a common reservoir for group I and group II introns (Hafez et al., 2013). Lichenized fungi (hereafter: lichens) are highly successful, widely distributed, and ecologically important (Hawksworth et al., 1984). Lichens consist of obligate symbioses between a minimum of one primary fungal partner (the mycobiont; often an Ascomycete) that provides structural protection for one or more primary photosynthetic partners (the photobiont; a green alga or cyanobacterium), which provide photosynthates to the mycobiont (Ahmadjian et al., 1981; Seaward, 1997; Brodo et al., 2001; Papazi et al., 2015). Lichens are particularly notable for their marked diversity in shape, size, and color, and have the ability to grow in nearly all terrestrial habitats, on many natural and artificial substrates (Fryday et al., 2007). Because the mycobiont provides physical protection to algal symbionts, the lichen symbiosis can survive in environments where neither partner could develop on its own, and thus this symbiosis contributes to the carbon cycle through the conversion of carbon dioxide to oxygen. Lichens also sequester compounds present in the environment including pollutants, and serve as important bioindicators for the health and history of myriad ecosystems (Szczepaniak and Biziuk, 2003).

As lichens are such an ideal study system we are going to look at intron movement in these species.

Previous work in plants and fungi have found that the cox1 gene is often parasitized by group I and group II introns (Cho et al., 1999; Hafez et al., 2013). The cox1 gene functions to…

We perform ancestral state reconstruction to determine the rate at which these parasitic elements are gained/lost from the cox1 gene. We have a time calibrated tree based upon fossil and molecular data and have used this to estimate the time of several discreet and continuous variables connected with intron polymorphisms.

Here we present genomic data from the obligate lichen symbiosis and use these data to test hypotheses of intron and retrotransposable element movement across the lichen tree of life. We examined 58 assembled and annotated mitochondrial genomes representing seven major clades of lichens and a morphologically diverse set of species from the southern Appalachian Mountains. We measure and compared the number of introns with and without associated intron encoded retrotransposable elements and perform ancestral state reconstruction to determine evolutionary timescale of loss and gain of these elements. Our results reveal differences in content and organization in lichen mycobiont mitochondrial genomes.

In Fungi: need more information for this! <https://www.sciencedirect.com/science/article/pii/S1087184513000194>

https://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-9-22

Questions:

**Understand the evolution of *cox1* introns in lichenized fungi. What was the ancestral state of a clade? How long do they last? Make a time tree like Sean. How quickly do they come in and out of the genome. Within a genus and do ancestral state reconstruction. Characteristic is number of introns in cox1.**

**Look up fossil records to calibrate nodes**

**Do introns contain evidence about their evolutionary history and their potential spreading strategy possibly involving homing endonucleases (i.e. can we tell the difference between horizontal and vertical transfer events?)**

**Can relate them to genome size, already shown, maybe check if there is a pattern of more introns/RTs in different growth forms?**

**Methods**

**Tissue Collection**

To test hypotheses of reductions in mitochondrial genomes of lichenized fungi, we selected 58 focal species that spanned a broad range of genetic distances present in the largest clade of lichens: the Lecanoromycetes (Miadlikowska et al., 2014). These taxa have diverse ecologies, thereby allowing us to explore whether mitochondrial genome evolution may be driven by a particular lifestyle, such as growth form or substrate. All 58 species were native to the study area and collected wild in the southern Appalachian Mountains during fieldwork in 2016. Museum voucher specimens were deposited at The University of Colorado (COLO) and The New York Botanical Garden (NY). Within 24 hours of initial field collection, we subsampled thalli as follows: for macrolichens, ca. 1 x 1 cm of tissue was removed, primarily from thallus margins and lobes. For microlichens, tissue was scraped from rock or tree substrates using a sterile razor blade; efforts were made to only sample one thallus for microlichens but given their physically small size, it is possible that multiple individuals per species were included in some of our sampling. Tissue samples removed for molecular study were air dried in a laminar flow hood for 24 hours then stored at -20°C until transport to COLO for extraction and sequencing.

**Genome Extraction and Sequencing**

Dried tissue samples were pulverized using tungsten carbide bearings in a Qiagen 96-well plate shaker. Genomic DNA was extracted from the lichen tissue samples using a Qiagen DNeasy 96 plant kit. The manufacturer’s protocol was modified to include a 10 minute 65°C incubation step for the ground material in lysis buffer as well as a 100% ethanol wash before final drying of the membrane prior to elution. Preliminary study found that these modifications improved lichen gDNA concentration and purity (Pogoda et al., unpublished data). Extracted samples were stored at -20°C prior to subsequent library preparation.

Genomic libraries were prepared using Nextera® XT DNA library prep kits (Illumina®) with the protocol optimized for 1 ng total of input DNA. Each gDNA sample was diluted to the appropriate concentration using a Qubit 3.0 fluorometer (ThermoFisher Scientific). Each sample was barcoded by the unique dual index adapters Nextera® i5 and i7. Resulting libraries were cleaned using solid phase reversible immobilization (SPRI) to remove fragment sizes less than 300 base pairs via an epMotion 5075TMX automated liquid handling system. Sample quality control (QC) was conducted to ensure appropriate sample concentration and fragment size using a Qubit 3.0 fluorometer and an Agilent 2100 Bioanalyzer prior to normalizing the loading concentration to 1.8-2.1pM with 1% PhiX control v3 added (Illumina®). Samples that passed QC were processed for paired end 150 base pair reads on the Illumina NextSeq® sequencer. All wet lab work was conducted in the Department of Ecology and Evolutionary Biology at the University of Colorado, Boulder. Sequencing was conducted at the University of Colorado BioFrontiers Institute Next-Generation Sequencing Facility in Boulder, Colorado.

**Mycobiont Genome Assembly**

To facilitate hypothesis testing, we focused on mycobiont mitochondria because of the availability of a comparative data set for non-lichenized mitochondrial genomes. All 58 mycobionts sampled in the present study represent Ascomycete fungi (i.e., no Basidiolichens were included). Raw de-multiplexed data for all 58 samples were sub-sampled to approximately 2GB per sample and trimmed using Trimmomatic-0.36 with the parameters “ILLUMINACLIP:NexteraPE-PE.fa:2:20:10MINLEN:140 LEADING:20 TRAILING:20” (Bolger et al., 2014) with the file “NexteraPE-PE.fa” containing the standard set of Nextera adapters to be trimmed from reads. Resulting fastq files were *de novo* assembled using SPAdes version 3.9 with the following parameters: SPAdes-3.9.0-Linux/bin/spades.py --careful -k 35,55,85 (Bankevich et al., 2012). The resulting assembly included genomic representatives of all taxa (e.g., primary mycobiont, secondary fungal partners such as endolichenic fungi and yeasts, bacterial symbionts) present in the metacommunity at the time of tissue sampling. Depth of the assembly is proportional to the amount of input DNA such that the primary fungal and photobiont partners typically have the highest coverage in contrast to other symbionts.

The following steps were taken to ensure the mitochondrial sequences presented here belonged to the mycobiont rather than the photobiont or any other symbiont (such as endolichenic fungi) present in the metacommunity at time of sampling, as follows. First, a command-line BLAST was conducted to identify candidate contigs as mitochondrial. Second, these contigs were then web BLASTed to the NCBI non-redundant database. In every taxon examined, the longest and highest-coverage contigs identified with the command-line BLAST had very high % identity web-BLAST hits to the expected lichenized fungus at common barcoding loci (Table XXX). We therefore concluded that these contigs belonged to the primary mycobiont mitochondrion as opposed to a more distantly related endolichenic fungus. These conclusions were borne out by the phylogenetic analyses described below. Third, contigs were circularized using the raw genomic reads and error-corrected using SAMtools *tview*, and *tview* was used to ensure that no contigs assembled as chimeras between the mycobiont mitochondrion and another mitochondrion present in the meta-assembly. Chimeric junctions appear as abrupt changes in alignment depth and sharp cutoffs in read alignments; *tview* revealed no chimerism in the assemblies.

To further validate the above method of identifying and assembling the primary mycobiont mitochondrial genome, we compared gene content, and overall genome size to four available primary mycobiont mitochondria downloaded from NCBI (*Lecanora strobilina*: KU308740; *Peltigera dolichorrhiza*: KT946595; *Peltigera malacea*: JN088164; and *Peltigera membranacea*: JN088165). Annotations were conducted using DOGMA (Wyman et al., 2004) and completed in Sequin 15.10 (Bethesda MD) using sequences from four genomes to confirm gene boundaries (*Lecanora strobilina*: KU308740; *Peltigera dolichorrhiza*: KT946595; *Peltigera malacea*: JN088164; and *Peltigera membranacea*: JN088165). The 58 lichen mitochondrial genomes were assembled and annotated by undergraduate and graduate students enrolled in the COLO Fall 2016 Genomics class taught by N. Kane, then checked for accuracy by the first author.

**Phylogenetic Comparative Analyses**

Using a concatenated sequence matrix composed of five conserved genes (*cox1, cox2, cox3, nad2,* and *nad4*) present in all 58 taxa, a best-estimate phylogenetic hypothesis was reconstructed using maximum likelihood (ML) methods implemented in RAxML v.8.0.0, with default parameters in effect (Stamatakis, 2014). We used this tree to first assess whether relationships among species reconstructed using our newly generated mitochondrial data yielded relationships consistent with prior studies; we compared our results to those inferred in Miadlikowska et al., 2014. Finally, to estimate the number of independent losses of a key gene of interest (*atp9*; see results), we conducted ancestral state reconstructions on the best estimate ML tree for presence vs. absence of this gene across sampled taxa. Analyses were conducted in Mesquite (Maddison and Maddison, 2017) using its ML ancestral state reconstruction package and default parameters.

Ahmadjian, V., & Jacobs, J.B. (1981). Relationship between fungus and alga in the lichen Cladonia cristatella Tuck. *Nature*, *289*(5794), 169-172.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., ... & Pyshkin, A.V. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*, *19*(5), 455-477.

Belfort, M., & Bonocora, R.P. (2014). Homing endonucleases: from genetic anomalies to programmable genomic clippers. *Homing Endonucleases: Methods and Protocols*, 1-26.

Bhattacharya, D., Friedl, T., & Helms, G. (2002). Vertical evolution and intragenic spread of lichen-fungal group I introns. *Journal of Molecular Evolution*, *55*(1), 74-84.

Bolger, A.M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170.

Brodo, I.M., Sharnoff, S.D., & Sharnoff, S. (2001). *Lichens of north America*. Yale University Press.

Chevalier, B.S., & Stoddard, B.L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic acids research*, *29*(18), 3757-3774.

Cordaux, R., & Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics*, *10*(10), 691.

Coros, C. J., Piazza, C. L., Chalamcharla, V. R., Smith, D., & Belfort, M. (2009). Global regulators orchestrate group II intron retromobility. *Molecular cell*, *34*(2), 250-256.

Edgell, D. R., Chalamcharla, V. R., & Belfort, M. (2011). Learning to live together: mutualism between self-splicing introns and their hosts. *BMC biology*, *9*(1), 22.

Fryday, A. M., Lendemer, J. C., & Howe, N. M. (2007). Porpidia soredizodes (lichenized ascomycota) in North America. *Opuscula Philolichenum*, *4*, 1-4.

Gray, M.W., Burger, G., & Lang, B.F. (1999). Mitochondrial evolution. *Science*, *283*(5407), 1476-1481.

Hafez, M., Majer, A., Sethuraman, J., Rudski, S. M., Michel, F., & Hausner, G. (2013). The mtDNA rns gene landscape in the Ophiostomatales and other fungal taxa: twintrons, introns, and intron-encoded proteins. *Fungal genetics and biology*, *53*, 71-83.

Haugen, P., Reeb, V., Lutzoni, F., & Bhattacharya, D. (2004). The evolution of homing endonuclease genes and group I introns in nuclear rDNA. *Molecular Biology and Evolution*, *21*(1), 129-140.

Hawksworth, D.L., Hill, D.J., & Phil, D. (1984). The lichen-forming fungi. Chapman & Hall.

Hausner, G. (2012). Introns, mobile elements, and plasmids. In *Organelle Genetics* (pp. 329-357). Springer Berlin Heidelberg.

Kanzi, A.M., Wingfield, B.D., Steenkamp, E.T., Naidoo, S., & van der Merwe, N.A. (2016). Intron Derived Size Polymorphism in the Mitochondrial Genomes of Closely Related Chrysoporthe Species. *PloS one*, *11*(6), e0156104.

Kennell, J. C., Moran, J. V., Perlman, P. S., Butow, R. A., & Lambowitz, A. M. (1993). Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. *Cell*, *73*(1), 133-146.

Lambowitz, A. M., & Zimmerly, S. (2004). Mobile group II introns. *Annu. Rev. Genet.*, *38*, 1-35.

Maddison, W.P. & Maddison D.R. (2017). Mesquite: a modular system for evolutionary anlaysis. Version 3.31 http://mesquiteproject.org

Mardanov, A.V., Beletsky, A.V., Kadnikov, V.V., Ignatov, A.N., & Ravin, N.V. (2014). The 203 kbp mitochondrial genome of the phytopathogenic fungus Sclerotinia borealis reveals multiple invasions of introns and genomic duplications. *PLoS One*, *9*(9), e107536.

Matsuura, M., Saldanha, R., Ma, H., Wank, H., Yang, J., Mohr, G., ... & Lambowitz, A. M. (1997). A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes & development*, *11*(21), 2910-2924.

Miadlikowska, J., Kauff, F., Högnabba, F., Oliver, J.C., Molnár, K., Fraker, E., ... & Otálora, M.A. (2014). A multigene phylogenetic synthesis for the class Lecanoromycetes (Ascomycota): 1307 fungi representing 1139 infrageneric taxa, 317 genera and 66 families. *Molecular Phylogenetics and Evolution*, 79, 132-168.

Michel, F., & Lang, B. F. (1985). Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. *Nature*, *316*(6029), 641.

Papazi, A., Kastanaki, E., Pirintsos, S., & Kotzabasis, K. (2015). Lichen symbiosis: Nature's high yielding machines for induced hydrogen production. *PloS one*, *10*(3), e0121325.

Printzen, C., & Ekman, S. (2003). Local population subdivision in the lichen Cladonia subcervicornis as revealed by mitochondrial cytochrome oxidase subunit 1 intron sequences. *Mycologia*, *95*(3), 399-406.

Rekadwad, B., Gonzalez, J. M., & Khobragade, C. N. (2017). Functional Diversity and Applications of Mobile Group II Introns. In *Drug Resistance in Bacteria, Fungi, Malaria, and Cancer* (pp. 161-169). Springer, Cham.

Robbins, J. B., Smith, D., & Belfort, M. (2011). Redox-responsive zinc finger fidelity switch in homing endonuclease and intron promiscuity in oxidative stress. *Current Biology*, *21*(3), 243-248.

Seaward, M.R.D. (1997). Major impacts made by lichens in biodeterioration processes. *International biodeterioration & biodegradation*, *40*(2-4), 269-273.

Sethuraman, J., Majer, A., Friedrich, N.C., Edgell, D.R., & Hausner, G. (2009). Genes within genes: multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of Ophiostoma and related taxa. *Molecular biology and evolution*, *26*(10), 2299-2315.

Schmelzer, C., & Schweyen, R. J. (1986). Self-splicing of group II introns in vitro: mapping of the branch point and mutational inhibition of lariat formation. *Cell*, *46*(4), 557-565.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313.

Thiéry, O., Börstler, B., Ineichen, K., & Redecker, D. (2010). Evolutionary dynamics of introns and homing endonuclease ORFs in a region of the large subunit of the mitochondrial rRNA in Glomus species (arbuscular mycorrhizal fungi, Glomeromycota). *Molecular Phylogenetics and Evolution*, *55*(2), 599-610.

Toor, N., Hausner, G., & Zimmerly, S. (2001). Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *Rna*, *7*(8), 1142-1152.

van der Veen, R., Arnberg, A. C., Van der Horst, G., Bonen, L., Tabak, H. F., & Grivell, L. A. (1986). Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell*, *44*(2), 225-234.

Wenzlau, J. M., Saldanha, R. J., Butow, R. A., & Perlman, P. S. (1989). A latent intron-encoded maturase is also an endonuclease needed for intron mobility. *Cell*, *56*(3), 421-430.

Xavier, B. B., Miao, V. P., Jónsson, Z. O., & Andrésson, Ó. S. (2012). Mitochondrial genomes from the lichenized fungi Peltigera membranacea and Peltigera malacea: features and phylogeny. *Fungal biology*, *116*(7), 802-814.