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. Probably need to talk about mitochondria. Maybe talk about the lifecyle of introns with ORFs, only if we want to look at fused/separate retrotransposable 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.

Homing endonucleases (HEs) are a class of rare cutting DNA enzymes. They are present in bacterial genomes as well as in the mitochondrial and chloroplast genomes of eukaryotes. These enzymes are found in group I and group II introns as well as inteins in their host genomes. There are at least six classes of HEs that are identified by unique amino acid sequences in their respective functional domains, LAGLIDADG, HNH, His-Cys box, GIY-YIG, PD-(D/E)xK and EDxHD; however the first four are the more common types. These enzymes have homing capabilities meaning that they are able to move themselves, as well as surrounding DNA, to a homologous allele that previously did not contain the element. Previous studies have shown that homing endonucleases play a role in size polymorphisms between mitochondrial genomes (Bhattacharya et al., 2002; Printzen & Ekman 2003; Haugen et al., 2004; Thiéry et al., 2010; Xavier et al., 2012; Beaudet et al., 2013; Mardanov et al., 2014; Kanzi et al., 2016). The homing endonucleases themselves are often small, less than 200 amino acids. These enzymes produce double stranded cuts at specific recognition sites, that can vary between 15-35 bp, followed by double stranded repair and insertion of an intron containing allele (Chevalier and Stoddard 2001; Thiéry et al., 2010). They can be inherited both vertically and horizontally and have been shown to transfer across biological kingdoms ((Cho et al. 1998; Goddard & Burt, 1999).

Homing endonucleases can either be encoded as a free standing intein or within an intron of another gene. The introns parasitized by HEs are group I and group II, and both have self-splicing capabilities. Group I introns and 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). Both types of introns have ribozyme activity. Group I introns have a conserved nine domains and group II have 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). It is thought that mobile group II introns were the predecessors 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). Homing of group I introns is initiated by the HE which upon recognizing a specific DNA sequence produces a double stranded break, where the intron is inserted and host machinery repairs the break using homologous recombination (Colleaux et al. 1986; Dujon, 1989; Belfort & Perlman, 1995). By contrast, group II intron homing is more complicated as it involves a reverse splicing reaction (Moran et al. 1995; Zimmerly et al. 1995a, b; Curcio & Belfort, 1996; Mills et al. 1996; Shearman et al. 1996; Guo et al. 1997; Matsuura et al. 1997; Cousineau et al. 1998; Yang et al. 1998). These self-splicing 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 I or 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). Much of the early work to understand these intron types and the HEs they encode was performed in fungi, specifically yeast. 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 and their retrotransposable elements (Hafez et al., 2013).

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 are often responsible for adding much of the vibrancy and texture to a given landscape. They have the ability to grow in nearly all terrestrial habitats, on many natural and artificial substrates, and are often the first species to colonize disturbed areas (Fryday et al., 2007). They contribute importantly to the process of soil formation, which is essential to the growth and development of plants and other organisms, by speeding the erosion of rocks and adding further nutrients to developing soils as a function of their own decomposition (Chen et al., 2000). 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 bioindicators for the health and history of myriad ecosystems (Szczepaniak and Biziuk, 2003). The lichenized fungal mitochondrion offers an important genetic resource to study evolutionary consequences of symbiosis.

Mitochondria are cellular organelles essential to eukaryotic life, as they provide ATP for use as energy in many necessary biosynthetic pathways (Gray et al., 1999). They are thought to be evolutionarily among the first landmark symbiotic events, having originated through endosymbiosis of an early alpha-proteobacterium into a host cell (Gray et al., 1999). Studies have demonstrated that the ancestral mitochondrion has undergone significant reductions in genome complexity and size via losses in protein-coding genes as well as rearrangements in gene order through processes similar to gene losses and rearrangements observed in more modern symbionts (Gray et al., 1999; Khachane et al., 2007). We sequenced 58 different species of lichen and assembled and annotated their mycobiont mitochondrial genomes. These 58 newly generated and annotated genomes represent seven major clades of lichens and a taxonomically diverse set of species from the southern Appalachian Mountains. We found widespread presence of introns that encoded retrotransposable elements, specifically homing endonucleases. Here we report intron sequence similarity across these species and discuss modes of inheritance (vertical and horizontal) of the retrotransposable elements. Additionally, we present a time calibrated tree of the 58 species examined and describe the ancestral states of for several variables (gene number, intron number, number of retrotransposable elements and number of empty introns). The *Usnea* clade in particular demonstrates a high degree of divergence and contained the highest number of introns but in contrast the fewest with homing endonuclease ORFs present.

References:

**Homing endonucleases from mobile group I introns: discovery to genome engineering (Stoddard)**

Structural conventions for group I introns (Burke)

Homing endonuclease structure and function (Stoddard)