

The phylogeography and invasion history of jumping worms on the Hamilton College campus
inferred through populations genetics

By Dylan Morse

Advisor: Professor Patrick Reynolds
Hamilton College

May 22nd, 2022

Table of Contents

Introduction pp. 4-14

- Importance
- Taxonomy and origin
- Morphological characteristics and life history
- Invasion history and ecological effects
- Phylogeography
- Molecular markers and DNA barcoding
- Genetic isolation, PCR, and sequencing
- Invasive species management
- Aims and hypotheses

Methods pp. 15-17

- Sample collection and preservation
- DNA extraction and amplification
- COI barcoding
- Haplotype analysis

Results pp. 18-22

- Population structure
- Haplotype analysis
- Haplotype lineage

Discussion pp. 23-25

- Amyntas* species as successful invaders of the Hamilton College campus
- Population structure and dynamics on the Hamilton College campus
- Haplotype diversity
- Haplotype lineages
- Introduction events

Conclusion pp. 26

- Larger importance
- Future research

Abstract:

Phylogeographic studies are helpful in reconstructing invasion history and population dynamics. Since the first introduction of *Amyntas* species into the United States in the 1930's, *Amyntas* species (invasive jumping worms) have spread throughout the eastern United States and into Canada. Three *Amyntas* species, *Amyntas agrestis*, *Amyntas tokioensis*, and *Metaphire hilgendorfi*, all affect productivity and nutrient cycling in soil and forested ecosystems. Recent studies have used the COI barcoding region to track invading taxa and elucidate cryptic diversity.

This study investigates populations of *Amyntas* species on the Hamilton College campus. Using the COI barcoding region to conduct a phylogeographic investigation, we analyzed (n = 80) specimens and found five distinct haplotypes on the Hamilton College campus. Three lineages were *Amyntas tokioensis*, the other two haplotypes were *Amyntas agrestis* and *Metaphire hilgendorfi*. The largest species sampled was *Amyntas tokioensis* (n = 63). *Amyntas agrestis* and *Metaphire hilgendorfi* had 11 and 5 individuals identified, respectively. Low haplotype diversity suggests a low rate of invasion and few introduction events. The large *Amyntas tokioensis* population and three haplotype lineages suggest that the species is more established in upstate New York. *Amyntas tokioensis* could also be outcompeting the other two species.

Introduction

Importance

Over the past two decades, the pace of worldwide industrialization and globalization has dramatically increased (Schult et al. 2016). The global connectivity of human populations has had wide-ranging effects on our communities and natural environments (Crown et al. 2008). New trade routes between previously disconnected and isolated countries, as well as enhanced transportation technology, have quickened the spread of non-native species (Schult et al. 2016). U.S. international trade and travel has provided pathways for foreign species transmission and introduction that were not possible a few decades ago (McCay et al. 2020). As global infrastructure continues to grow, human-mediated transportation pathways, such as airplanes or cargo barges, have further accelerated the rate of dispersal of non-native species (Schult et al. 2016; McCay et al. 2020). This is because invasive species are likely to spread through pathways that have high levels of human activity (Cameron & Bayne 2009; Novo et al. 2015). These pathways have large dispersal corridors that are permeable to invasion and are regularly disturbed, making them susceptible to opportunistic fast-growing species (Cameron & Bayne 2009; Ruis and Turon 2012). The trans-oceanic transportation of invasive species can occur in the ballast water of ships, in organic materials, or with plants and produce (Cameron et al. 2008; Schult et al. 2016; Chang et al. 2016).

Corn et al. 1999 estimated that of the 50,000 species in the United States, 4,300 were non-native and invasive species. While invasive species stem from many geographic regions around the world, an increasing number of invasive species are coming from Asia (Crown et al. 2008). Numerous non-native species from Asia, such as carp, mussels, and earthworms, have successfully implanted themselves within North American ecosystems and have caused great ecological and economic damage (Crown et al. 2008). Here in New York State, species such as the Asian Longhorn Beetle and Hydrilla, both which originate around Australia and Asia, have recently had devastating effects on terrestrial and aquatic environments (DEC). In fact, invasive species are the second leading cause, after human population growth, of species and population extinction and endangerment in the United States (Crown et al. 2008). Invasive species alter ecological dynamics, shift richness and abundance, and drive extinction (Cameron et al. 2008; Crown et al. 2008; Chang et al. 2016; Schult et al. 2016). Annually, invasive species cost the U.S.

an estimated 120 billion dollars in environmental damage and restoration efforts (Crowl et al. 2008).

While conservationists and scientists recognize the importance of understanding how invasive species affect different ecosystems, little attention has been devoted to invasive soil-dwelling organisms despite their importance to soil ecosystems (Novo et al. 2015). Invasive Asian earthworms, in the genus *Amyntas* and family Megascolecidae, have been of particular concern because of the harmful effects they have on soil ecosystems, causing significant changes in forest ecosystems (Schult et al. 2016). This serves to orient the reader to Asian invasive earthworms. Specifically, three species of Asian invasive earthworms, referred to as invasive jumping worms, common in New York (McCay et al. 2020; Chang et al. 2021). These three species, *Amyntas agrestis*, *Amyntas tokioensis*, and *Metaphire hilgendorfi*, all have established populations locally (Chang et al. 2021).

Taxonomy and origin

Amyntas agrestis, *Amyntas tokioensis*, and *Metaphire hilgendorfi* are referred to as pheretimoids (McCay et al. 2020). Pheretimoids belong to the Megascolecidae family in the *Pheretima* complex (Chang et al. 2016; McCay et al. 2020; Chang et al. 2021), this complex includes members of *Amyntas*, *Metaphire*, *Pheretima* species, and related genera. Worldwide, soil organisms make up around one fourth of described biotic life, earthworms comprise about 3,700 species (Dacaens et al. 2013). There are more than 1000 described species of pheretimoids worldwide, and the species-rich group is dominant in forest, grassland, and agricultural communities around the Asian-Pacific (Chang et al. 2021). *Amyntas* and *Metaphire* originate in tropical and temperate East Asia, mainland Southeast Asian, and Australia (McCay et al. 2020; Chang et al. 2021). However, pheretimoids have become abundant in South Africa, South America, Oceania, and North America (McCay et al. 2020). Many studies have referred to their spread as a “second earthworm diaspora” (McCay et al. 2020). There are 16 pheretimoid earthworm species currently known to have invaded North America (Chang et al. 2021).

Invasive jumping worms are now widespread in the Eastern United States. *Amyntas agrestis*, *Amyntas tokioensis*, and *Metaphire hilgendorfi* have been identified throughout the Southeast, Mid Atlantic, and Northeast (Schult et al. 2016; Chang et al. 2016). Chang et al. 2016 identified the range of *Amyntas agrestis*, *Amyntas tokioensis*, and *Metaphire hilgendorfi* to

extend as far west as Wisconsin and Illinois and as far north as Maine. They have not yet been documented in the U.S. west of the Rocky Mountains (Chang et al 2021). This may be because habitat matching facilitates invasion, the similarity of subtropical and temperate climates in Asian countries helped *Amyntas* first colonize the southern U.S (Novo et al. 2015). Locally in Clinton, New York, studies from the past decade have identified *Amyntas agrestis* and *Amyntas tokioensis* (Bohlen et al. 2004; Schult et al. 2016; Chang et al 2021). However, research in the Reynolds lab at Hamilton College have sampled all three species, including *Metaphire hilgendorfi* which wasn't shown to be present before.

Morphological characteristics and life history

Pheretimoids have common morphological characteristics (McCay et al. 2020). These three species look very similar and, because they live in the same habitat, they are all often mistakenly identified as *Amyntas agrestis* (Chang et al. 2016; Chang et al. 2018). In general, earthworms are identified by their internal and external reproductive organs, length, color, segmentation, and size (Chang et al. 2021). Because of this, identifying juveniles by morphological characteristics to the species level is not possible, in adults it is still difficult to identify individual specimens to the species level because of male reproductive degeneration and parthenogenesis (Chang et al. 2021). They have long bodies and cocoons that are resilient and long lived (McCay et al. 2020). This, combined with the fact they have broad diets and face low predation and parasitism rates, allows them to rapidly spread and colonize new environments (McCay et al. 2020).

In North America, there are several regional/common names for Asian invasive earthworms, these earthworms are named for behavioral characteristics (Chang et al. 2021). In Kansas, invasive Asian earthworms are called disco-worms but in a majority of the United States they are known as jumping worms (Chang et al. 2021). Invasive jumping worms earned their name because of their serpentine motion and ability to jump on surface soils. Not all pheretimoids jump, however that jumping motion is an unintended consequence of their violent thrashing form of locomotion (Chang et al. 2021).

Amyntas agrestis ranges from 70-160 mm by 5-8 mm in size and usually has 63-110 segments. Specimens are red, yet the species is difficult to identify because of the lack of post-clitellar genitalia. *Amyntas agrestis* is epi-endogeic and has invaded U.S. forests because of its

dietary flexibility. The species competes for local niches with species such as millipedes and is commonly used in commercial mulch for horticulture and landscaping, which most likely has quickened their spread. *Amyntas tokioensis* is slightly smaller, ranging from 30-125 mm by 3-7 mm in size. The species has 84-102 segments that are colored red or brown. The species has post clitellar genitalia markings with small, circular disks on each side. The species was historically referred to as *Pheretima levis* or *Metaphire levis* and is believed to have been directly introduced from Japan in 1947 (Chang et al. 2021). Reproduction of *Amyntas tokioensis* is parthenogenetic. *Metaphire hilgendorfi* was first known as *Amyntas hilgendorfi* in North America. The species is 109-170 mm by 6-8 mm in size and has 98-118 segments. Usually, specimens are red or reddish brown when alive. *Metaphire hilgendorfi* is only one of two species present in the United States that has spread far north into Canada (All from Chang et al. 2016).

The three species share an annual life cycle. Cocoons overwinter from the end of November to April (McCay et al. 2020). Cocoons are especially hearty and resilient; they extend propagules that help them endure tough winter conditions or overpopulation waiting for the right time to hatch (McCay et al. 2020). In May and June juveniles hatch and mature, from July to October adults prepare for mating (McCay et al. 2020).

Invasion history and ecological effects

Native earthworms were exterminated from most of North America during the Pleistocene glaciations more than ten thousand years ago (Cameron et al. 2008; Schult et al. 2016; Chang et al. 2016). As the Wisconsin glacial shield retreated, few native earthworms were left. European earthworms were introduced to the Americas during European colonization, most likely introduced through the dumping of materials used in ballast and transport of goods and materials on ships (Cameron et al. 2008; Schult et al. 2016; Chang et al. 2016). Today, the dominant earthworm species currently inhabiting North America are native to Europe (Schult et al. 2016). While European earthworms have had significant ecological impacts on Northern ecosystems, *Amyntas* are of special concern because their environmental effects and range are not fully understood (Schult et al. 2016).

Amyntas agrestis, *Amyntas tokioensis*, and *Metaphire hilgendorfi* are common throughout the eastern United States despite their relatively recent introduction and highly damaging to forested ecosystems (Chang et al. 2021). *Amyntas agrestis*, *Amyntas tokioensis*,

and *Metaphire hilgendorfi* were not present within the United States until the 1930's. *Amyntas agrestis* was first sampled in 1939 in Maryland. *Amyntas tokioensis* was first sampled in 1947 in New York City. *Metaphire hilgendorfi* was first sampled in 1948 in Albany, New York (Chang et al. 2021). Some scientists believe that these species arrived with cherry trees donated by Japan to the United States that were planted around Bethesda, Maryland and Washington D.C. (Chang et al. 2021). However, presumably invasive jumping worms had more than one introduction event as U.S. trade increased with the Pacific over the last century (Schult et al. 2016). Once within the United States, invasive earthworms most likely spread along commonly used transportation routes such as highways (Cameron et al. 2008). *Amyntas* have especially high rates of mobility, reproduction, and are epigeic, which allows species to be opportunistic and respond to environmental disturbances that other species cannot (Novo et al. 2015). Notably, *Amyntas*' ability to aestivate, its physiological plasticity, and parthenogenesis could also aid in invasion (Novo et al. 2015).

Asian invasive jumping worms have wide-ranging effects on the abiotic and biotic environment (Bohlen et al. 2004; Novo et al. 2015; Laushman et al. 2018). Earthworms are ecosystem engineers, causing significant changes to the structure and functioning of the ecosystems they inhabit by burrowing, mixing, and casting soils (Vitousek 1990; Bohlen et al. 2004; Cameron et al. 2008). Ecosystem engineers are defined as organisms that directly or indirectly modulate the availability of biotic and abiotic resources (Vitousek 1990). Earthworms are specifically allogenic engineers because they change biotic and abiotic materials from one state to another (Vitousek 1990). This is opposed to autogenic ecosystem engineers that change the environment through their own physical structures, an example of which are beavers (Vitousek 1990). Earthworms change the ecosystem, specifically soils, through altering the physical environment and limiting the quality and quantity of resources available to native species (Cameron et al. 2008; Novo et al. 2015; Schult et al. 2016). Because earthworms modify the surface soil, which is essential for plant and fungal growth (Bohlen et al. 2004; Laushman et al. 2018). Therefore, earthworm invasions have significant impacts on nutrient cycling, food webs, and plant communities that depend on the exchange of resources and energy between the soil and upper-level ecosystems (Bohlen et al. 2004; Laushman et al. 2018). Earthworm invasions can rapidly alter soil structure, humus forms, and plant communities (McLean & Parkinson 2000; Bohlen et al. 2004; Schult et al. 2016; Laushman et al. 2018). This occurs

because earthworms consume some of the carbon stored in leaf litter and redistribute it much more quickly than normal environmental processes (Laushman et al. 2018). In addition, their burrowing and feeding behavior redistributes nutrients and changes the properties of soils and soil pH (McLean & Parkinson 2000; Bohlen et al. 2004). Earthworms shift the soil cycling from a slower cycling system, dominated by fungi, to a faster cycling system, dominated by bacteria (Bohlen et al. 2004). Dramatic changes in litter-soil composition and habitat have cascading effects in the organisms living in it (Bohlen et al. 2004; Cameron et al. 2008; Schult et al. 2016). Organic material within soils provides microhabitats for microbial and fungi communities (McLean & Parkinson 2000). Due to the feeding behavior and effects of earthworms on soil composition, disturbed soils tend to favor opportunistic fast-growing fungi and limit competition (McLean & Parkinson 2000; Bohlen et al. 2004). Some studies have found that earthworm invasions alter local microbial and fungi communities (McLean & Parkinson 2000). In some cases, the effects of epigeic invasive earthworms correlated positively with fungal dominance and negatively with fungal richness and diversity (McLean & Parkinson 2000).

Invasive earthworms can have detrimental effects on vertebrates as well. Earthworms are important in Red-backed salamander diets (Bohlen et al. 2004; Student & Marchand n.d.). Salamanders eat earthworms which are a protein rich food. The presence of earthworms in adult salamander diets has been seen to increase adult salamander health and fecundity (Bohlen et al. 2004; Student & Marchand n.d.). However, some studies have found that there is a negative correlation between the presence of earthworms and juvenile salamander health (Bohlen et al. 2004; Student & Marchand n.d.).

Phylogeography

Assessing, researching, and addressing the effects of invasive jumping earthworms requires a comprehensive understanding of species and population distribution. Multiple methodological processes have been used to identify species and track their distribution throughout the United States. Phylogeography is an integrative discipline that attempts to understand the geographic and spatial ordination of genotypes (Rius & Turon 2012; Novo et al. 2015). Phylogeography aims to investigate the relationship among geographic history and the mechanisms driving speciation (Rius & Turon 2012). Phylogeographic analyses have been used to create haplotype networks, mismatch analyses, genetic differentiation estimators, and

phylogenies (Ruis and Turon 2012). This approach has become important in the study of invasive and non-native species because phylogeography can be used to reconstruct the history of species invasion and can help in predicting or managing the spread of invasive species (Schult et al. 2016). Phylogeographic analysis can determine if non-native populations are founded by a single or multiple introduction event (Cameron et al. 2008). Populations introduced through a single introduction event tend to show reduced genetic diversity and variation compared to populations founded through multiple introductions events (Cameron et al. 2008). Single introduction events result in genetic bottlenecks that lead to reduced genetic diversity (Novo et al. 2015). Because founding events will dramatically decrease genetic variation within new populations, invasive populations should be less genetically diverse than the populations which they come from (Ruis and Turon 2012; Novo et al. 2015). However, there are various other mechanisms that can affect and reduce genetic variation, such as reproductive isolation due to mutation or hybridization (Ruis and Turon 2012). Some founding populations do not experience reductions in genetic diversity because their fixed alleles confer reproductive success potential (Ruis and Turon 2012). Multiple introduction events promote successful colonization and invasion because they enhance genetic diversity and, subsequently, suggest the rapid adaptation and expansion of a species range (Cameron et al. 2008; Novo et al. 2015).

Phylogeographic analysis hinges on a healthy sample size of specimens and assumes long-term natural evolutionary processes occur (Ruis and Turon 2012; Novo et al. 2015). The term should only be used to refer to specific analytical methods used to infer historical processes leading to modern geographic distribution (Ruis and Turon 2012). Phylogeography, being hypothesis driven, also depends on certain assumptions about mutation, drift, and migration being at equilibrium (Ruis and Turon 2012; Chang et al. 2021). There are also unique benefits of phylogeography, each specimen within a study can be individually identified, and molecular methods/barcodes of identification are available for pheretimoid earthworms.

Molecular markers and DNA barcoding

An essential discovery in evolution biology and ecology is the discovery of molecular markers (Dai et al. 2012). The use of molecular markers and DNA barcoding is revolutionary because it allows for the identification of specimens when morphological identification isn't possible. Earthworm identification at the species level requires high-level taxonomic expertise

and is prone to error (Decaëns et al. 2013). This is because earthworms lack common, stable diagnostic characteristics (Decaëns et al. 2013). Juveniles also commonly lack many of the sexual attributes that can be used to identify adults and other traits can be environmentally induced (Decaëns et al. 2013). DNA barcoding has been tied to not only in species diversity research, but has been essential in cryptic diversity, alpha taxonomy, and phylogeography and population genetics (Decaëns et al. 2013). While DNA barcoding most likely can't help with supra-specific species level identification and analysis, the phylogeographic information it contains is beneficial at the species and community level (Decaëns et al. 2013). Notably, one of the first uses of DNA barcoding in the phylogeographic analysis of earthworms was the Chang and Chen 2005 study that used COI patterns to propose the biogeographical history of *Amyntas* species in Taiwan. In our research, DNA barcoding could be specifically used in the phylogeographical study of distribution, nucleotide, and haplotype diversity to recreate and interpret the history of local *Amyntas* invasions.

Many scientists have argued that the COI barcoding region should be a global standard for identification (Hajibabaei et al. 2007, Decaëns et al. 2013, and Nouri-Aiin et al. 2021). The mitochondrial cytochrome c oxidase subunit I (COI) has been used as a standard genetic marker, or barcode, for animals because of its stability (Cameron et al. 2008; Dai et al. 2012; Schult et al. 2016; Chang et al. 2021). The 5' end of the COI gene has proven less effective with diverse taxa such as algae and fungi (Dai et al. 2012). Studies have suggested using various parts of the mtDNA to look for other barcode loci (Dai et al. 2012). The nuclear ITS regions ITS1 and ITS2 have been used for barcoding in fungi and plants (Dai et al. 2012). DNA barcoding is based on the premise that individuals within a species have distinct mutations within the genome (Hajibabaei et al. 2007). Specifically, variation within these short genetic sequences allow researchers to determine genetic and phylogenetic relations because genetic variation between species exceeds that within species (Hajibabaei et al. 2007).

Like the barcode itself, methods of analysis are wide ranging depending on each study's intended purpose. Phylogeographic models can be used to characterize the dynamics of invasive species spread (Lustig et al. 2017). Examples of methods include Bayesian methods, clustering methods, tree-based models, and distance-based models (Dai et al. 2012). Analyses of nuclear and mitochondrial genetic material can reveal origins and phylogeography of invasive species (McCay et al. 2020). Specific models can even account for the role of landscape structure,

dispersal patterns, and survival and reproduction rates within a population (Lustig et al. 2017). Genomic data is invaluable in addressing taxonomic mysteries. Phylogeographic analyses are most used to detect patterns of evolutionary change over time and space. However, phylogeographic analyses can also be used to track cryptic invasions, species origin and pathways, invasion events, and assess hybridization and introgression (Ruis and Turon 2012). Well established phylogeographic analyses, such as FST and AMOVA, can be based on changes in allele frequency. Bayesian models and highly advanced GTR+I+G evolutionary are also commonly used in phylogeographic analysis (Ruis and Turon 2012; Schult et al. 2016). These analyses are essential in the management and control of invasive species.

Genetic isolation, PCR, and sequencing

The genetic identification of individual specimens, specifically of the COI gene, follows well understood protocols (Cristescu 2015). Samples used for genetic identification are usually snippets of tissue or tail snips. Total genomic DNA is extracted from tissue using DNA isolation and purification kits such as the DNeasy Blood & Tissue Kit (Cristescu 2015; Nouri-Aiin et al. 2021). Isolated gDNA is then washed, purified, and cleaned for PCR amplification (Nouri-Aiin et al. 2021). The ~650 base pair fragment of the COI gene is amplified using kits like the HotStartaq Plus mm Kit; COI primers are well known, and PCR products usually need to be visualized on agarose gels before they are sent to laboratories for exact genomic sequencing. Some species can be identified using much shorter segments of the COI gene, especially in samples that have been degraded (Hajibabaei et al. 2007). PCR is usually conducted using MegaCOI-F and MegaCOI-R primers (Cristescu 2015). There are distinct PCR protocols for amplifying the COI and ITS regions. The COI PCR is 94°C for 2 minutes, 40 cycles of 94°C for 20 seconds, 54°C for 20 seconds, 72°C for 45 seconds, and a final extension at 72°C for 10 minutes (Dai et al. 2012). Sequencing is usually done by an outside laboratory and sent back to scientists for phylogeographic analysis (Cristescu 2015). Once genetic sequences are identified they are compared to known earthworm sequences using the BLAST function in GenBank.

Many molecular markers and analytical approaches have been employed to reconstruct invasion histories from genetic information (Schult et al. 2016; Decaëns et al. 2013; Dai et al. 2012). However, there are next generation techniques that can be used for DNA extraction, amplification, and sequencing that aren't as labor intensive and costly as traditional DNA

barcoding. Nouri-Ajin et al. 2021 developed a method that can distinguish between species of concern using simple gel electrophoresis of amplified PCR products. Specifically, the multiplex protocol targets differences within the COI gene of *Amyntas agrestis*, *Amyntas tokioensis*, and *Metaphire hilgendorfi*. The protocol identified differences within the COI gene of each species to yield different amplicon sizes. Thus, after isolating genetic material PCR is conducted with species specific primers that will replicate regions of distinct sizes. This allows the use of simple gel electrophoresis to differentiate between *Amyntas agrestis*, *Amyntas tokioensis*, and *Metaphire hilgendorfi*. The method costs a fraction of the cost of COI identification and sequencing because it doesn't involve sending products to a separate lab, and it has been proven to be 100% effective and accurate with all three *Amyntas* species discussed here (Chang et al 2021; Nouri-Ajin et al. 2021).

Invasive species management

The control of invasive species has become one of the most expensive and urgent tasks in the U.S. (Vitousek 1990; Crawl et al. 2008). Invasive species have profound consequences on the ecosystems they occupy (Crawl et al. 2008). While *Amyntas* invasions may not have disastrous effects on forest ecosystems, the disturbances they cause make secondary invasions more likely (Vitousek 1990). Preservation and conservation measures have traditionally worked to protect natural communities from disturbances (Vitousek 1990). However, modern empirical evidence and theory has demonstrated moderate levels of disturbances promote maximum species richness (Vitousek 1990). Therefore, conservation and preservation plans addressing *Amyntas* invasions should look to balance disturbances caused by Asian invasive earthworms with ecological disturbances, such as fires, storms, and droughts, important in maintaining ecosystem health.

Amyntas management and control plans have developed in complexity over the past decade. Early detection is essential for stopping the future spread of invasive organisms. The use of citizen science, awareness, and hotline programs all help in tracking the spread of invasive species in real time (McCay et al. 2020). Master gardeners and composters, specifically, are groups that are highly trained in identifying earthworm species and are likely to come in contact with them regularly (McCay et al. 2020). EMG hotline exists for reporting invasive specimens and programs such as WormWatch and Open-Air Laboratories are engaging citizen scientists in

soil reporting and monitoring programs throughout the eastern U.S. and United Kingdom (McCay et al. 2020). Because pheretimoids, such as *Amyntas agrestis*, *Amyntas tokioensis* and *Metaphire hilgendorfi*, are epi-endogeic and spend time at the soil surface, researchers have looked at if physical control is possible. Hand collection does not seem practical because many adults will burrow. Prescribed fires have been used to treat infected forests, because cocoons cannot survive temperatures above 40° C. This is one possible option for treatment (McCay et al. 2020). While studies have found *Amyntas* specimens are killed by prescribed fires, most research found no significant difference in populations between experimental burn-plots and control plots (McCay et al. 2020). Identifying thermal limits of *Amyntas* cocoons, juveniles, and adults could create novel ways to treat them using composting and solarization. Chemical control is also difficult because there are no known pesticides currently (McCay et al. 2020).

Aims and hypothesis

This study intends to look at the species history, geographic origin, and spread, of Asian invasive earthworms. We investigate the history of these earthworms' invasions into the local upstate New York region, and how phylogeography can be used to study their invasion history.

The aim of this study is to use previously captured *Amyntas agrestis*, *Amyntas tokioensis*, and *Metaphire hilgendorfi* to explore the genetic diversity and phylogenetic relationships of each species within local populations on the Hamilton College campus. Furthermore, using genetic and phylogenetic relatedness we aim to investigate the history of jumping worms' invasions into the local upstate, New York region.

Methods

Sample collection and preservation

Amyntas samples were collected in September and October of 2020 on the Hamilton College campus. Earthworms were sampled at distances of 0, 10, 20, 30, and 40 meters from a nearby path in the Root Glen near Bundy Café at five-minute search intervals. Samples (n = 94) were extracted manually by hand. After collection, worms were anesthetized in 10% ethanol. A tail snip of 1–2 mm of tissue was taken and preserved in RNAlater for total genomic DNA isolation in the Reynolds Lab at Hamilton College.

DNA extraction and amplification

For DNA extraction, amplification, and sequencing, tail snips were removed from RNAlater and digested with proteinase K and gentle agitation for 12 hours at 56°C. Once tissue was broken down, total genomic DNA was isolated via elution using the QIAGEN (Germantown, MD; www.qiagen.com) DNeasy Blood & Tissue Kit (Appendix 1). The gDNA was labeled and stored at -20°C.

The COI (cytochrome c oxidase subunit I) gene was chosen for amplification and specimen identification as described above (Hajibabaei et al. 2007; Decaëns et al. 2013; Nouri-Aiin et al. 2021). MEGA primer sequences, designed specifically for *Amyntas agrestis*, *Amyntas tokioensis*, and *Metaphire hilgendorfi* (Schult et al. 2016), were used for polymerase chain reaction (PCR) amplification, using QIAGEN HotStarTaq DNA polymerase (Appendix 2). MegaCOI-F primers were [5'-TAYTCWACWAAYCAYAAAGAYATTGG-3']. MegaCOI-R primers were [5'-TAKACTTCTGGRTGMCCAAARAATCA-3'].

Forty PCR cycles of denaturation at 95°C, annealing at 40°C, and extension at 72°C were preceded by 95°C for 5 minutes to activate the DNA polymerase and followed by 5 mins at 72°C and stored until collection at 4°C and stored long-term at -20°C. PCR amplicons were visualized on 1% Agarose with 0.5X TBE buffer (Appendix 3). Agarose gels were imaged on Bio-Rad Chemidoc Xrs+ System using UV trans illumination to check for amplification success and purity before sequencing (Appendix 4).

COI barcoding

COI barcoding of the specimen samples collected ($n = 94$), 54 specimen samples were processed in academic year 2020-21 and 40 were processed during academic year 2021-22 (by the author and Kiana Arcayena). Amplicons were sent to GENEWIZ for sanger sequencing (South Plainfield, NJ; www.genewiz.com) (Appendix 6). Sequences were edited in Unipro's UGene. Consensus sequences were constructed from forward and reverse contiguous sequences using MUSCLE alignment. Species identification was done using NCBI Nucleotide Blast (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), based on percent identity and E-value (Appendix 7).

Sequences PCR# 29.6, 29.8, 30.4, 30.5, 30.6, 30.7 and 30.8 were omitted because of errors in sanger sequencing. Previously, seven sequences were omitted from analysis during analysis conducted during the 2020-2021 academic year.

Consensus sequence alignment and neighbor-joining tree construction, to confirm species identification, was completed using UGENE (<http://ugene.net>). Consensus sequences from *Amyntas* samples were aligned using MUSCLE (Appendix 8). Aligned sequences were trimmed and a neighbor joining tree was created using PHYLIP neighbor joining and the Kimura distance matrix model (Appendix 8).

Haplotype analysis

Haplotype analysis was conducted to determine genetic structure of the populations present. Unique haplotypes and haplotype diversity were analyzed using FASTA Tools Unique Sequences Program from NCBI (www.ncbi.nlm.nih.gov/CBBresearch/Spouge/html_ncbi/html/fasta/uniqueseq.cgi). Haplotype relationships and nucleotide diversity were analyzed using PopART (<http://popart.otago.ac.nz/index.shtml>). Haplotype relationships were discovered using a minimum spanning network model. Files needed for FASTA and PopART programs were edited in TextEdit. Polymorphic sites as well as uncorrected pairwise p-distances among lineages and within lineages were considered using MEGA (Appendix 9). Each haplotype was searched for in NCBI Blast to establish at what other locations each has been sampled. Each haplotype was identified as the same by a 100% query match. Accession numbers were paired with studies and the geographic locations of where each five haplotypes have been sampled was noted.

The alignment and reading of base specimen sequences was completed four times to ensure accuracy. Deviations in the number of distinct haplotypes identified by UGENE, FASTA,

and PopART led us to reanalyze data. Through three primary analyses, errors in sanger sequencing reads could have affected the accuracy of sequence alignment and haplotype analysis. Some sequences had read errors due to high static or low signal clarity and so did not align with high accuracy. These sequences were isolated and read errors were identified using sanger sequencing electropherograms to identify nucleotide bases. After clarifying nucleotide bases, sequences were realigned and analyzed in UGENE, FASTA, and PopART.

Results

Population structure

In total, 80 of 94 collected specimens were sequenced and identified as *Amyntas* species. There were 63 *Amyntas tokioensis*, 11 *Amyntas agrestis*, and 5 *Metaphire hilgendorfi* specimens (Fig. 1), by percentage 78.8%, 13.9%, and 6.3%, respectively. Species identifications from NCBI Blast were corroborated using a neighbor joining tree, a bottom-up clustering method, to identify distinct lineages and species (Fig. 2). Of the 80 specimens, 60 were from the Bundy Café site and 20 were from the Root Glen site. At Bundy Café, there were 48 *Amyntas tokioensis*, 11 *Amyntas agrestis*, and 1 *Metaphire hilgendorfi* specimens; at the Root Glen site, 15 *Amyntas tokioensis*, 0 *Amyntas agrestis*, and 5 *Metaphire hilgendorfi* specimens (Fig. 3).

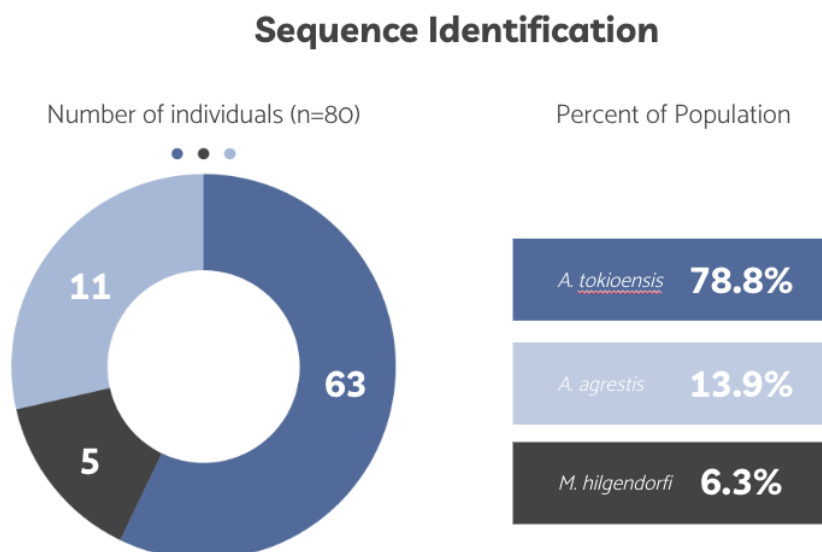


Figure 1 shows the number of specimens ($n = 80$) identified as *Amyntas tokioensis*, *Amyntas agrestis*, and *Metaphire hilgendorfi* on the left. On the right, the *Amyntas* species population is broken down into percentages.

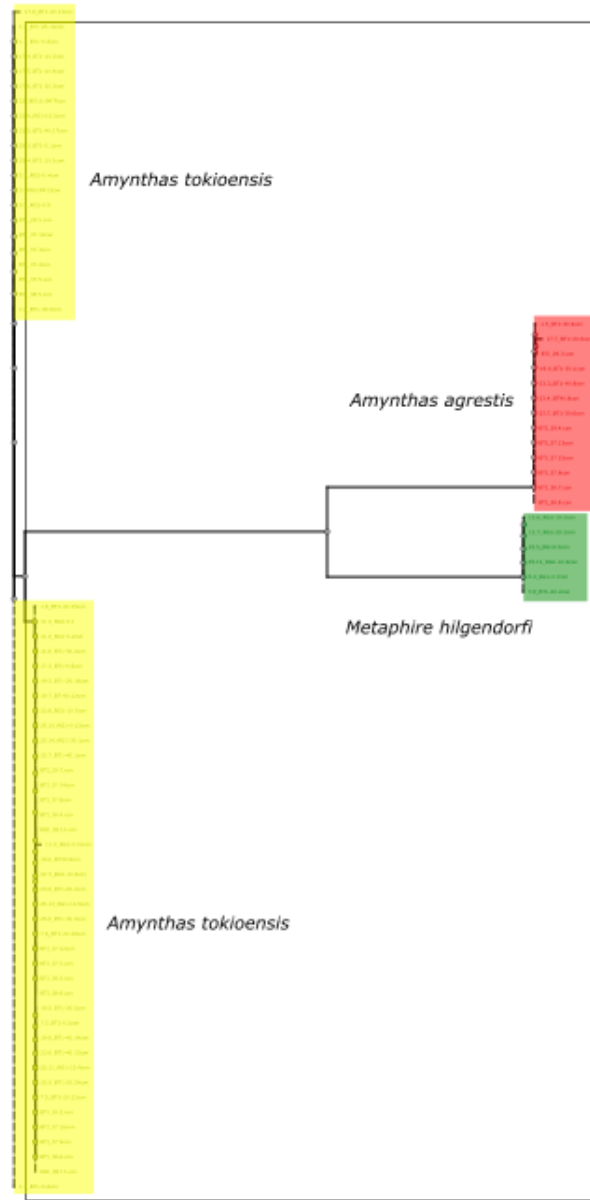


Figure 2 shows the phylogenetic neighbor joining tree created in UGENE for sequenced specimens ($n = 80$). *Amynthus tokioensis* specimens are highlighted in yellow on the left. *Amynthus agrestis* are grouped in red and *Metaphire hilgendorfi* is grouped in green on the right.

Haplotype analysis

Five unique haplotypes were identified. The minimum spanning tree illustrates relatedness and the lineages of *Amynthus* species present (Fig. 4). Each distinct haplotype is identified by species: *Amynthus tokioensis* 1 (T1), *Amynthus tokioensis* 2 (T2), *Amynthus tokioensis* 3 (T3), *Amynthus agrestis* 1 (A1), and *Metaphire hilgendorfi* 1 (H1) (Fig. 5). All three

Amyntas tokioensis haplotypes were present at both Bundy Café and Root Glen sites, as was the single haplotype of *Metaphire hilgendorfi*. The single *Amyntas agrestis* haplotype was present only at Bundy Café.

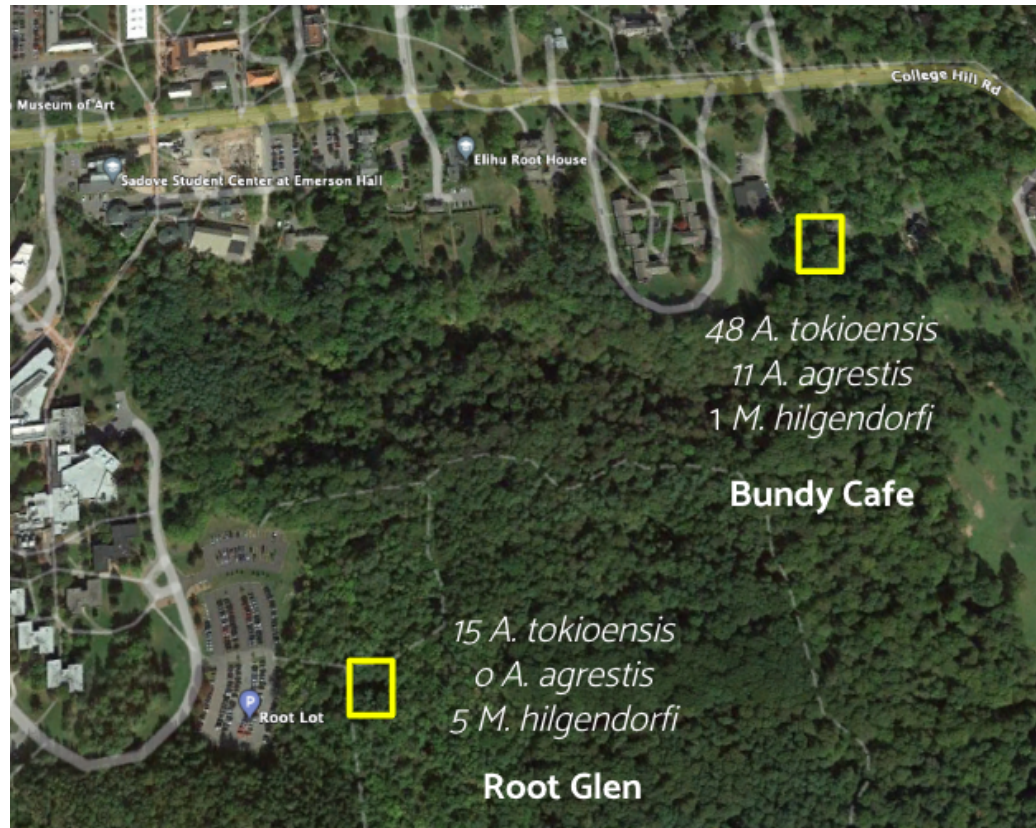


Figure 3 shows the number of *Amyntas* species identified at each study site. The study sites are highlighted in yellow. The majority of specimens came from Bundy Café and *Amyntas tokioensis* dominates both populations. No *Amyntas agrestis* were found at the Root Glen study site.

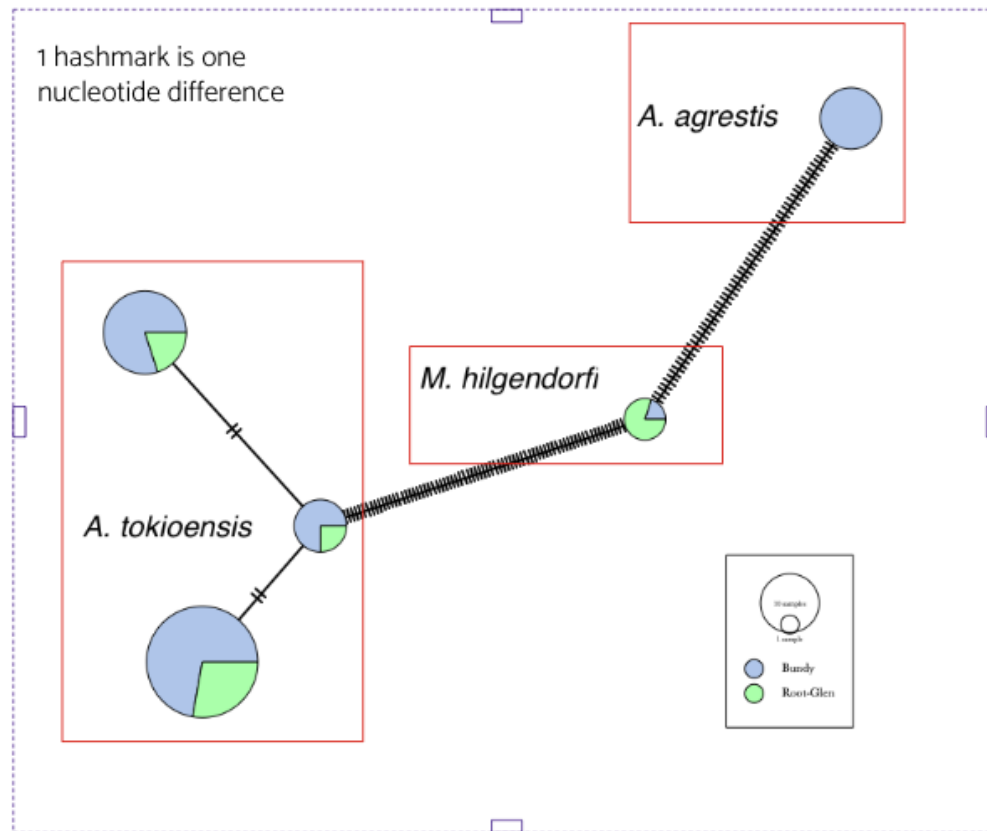


Figure 4 shows the minimum spanning tree created in PopART for all specimens ($n = 80$). Haplotypes are divided from each other by hash marks. One hashmark is one nucleotide difference. Specimens from the Bundy Café site are shown in blue, while specimens found at the Root Glen site are green.

Haplotype lineages

Haplotype T1 had 20 individuals and was identified to be *Amyntas tokioensis*. Using NCBI Blast results, the haplotype lineage was identified internationally in Japan and in the United States in Arkansas and Missouri. Haplotype A1 had 11 individuals and was identified to be *Amyntas agrestis*. NCBI Blast results showed that the haplotype was found in Japan and the South Kuril Islands of Russia. Haplotype T2 had 8 individuals and was *Amyntas tokioensis*. NCBI Blast results showed that the haplotype was identified in Japan, the South Kuril Islands, and the northeast United States. Haplotype T3, the largest lineage, had 35 individuals and was *Amyntas tokioensis*. This haplotype was found internationally in Japan, the South Kuril Islands, and the northeast United States. More specifically, Schult et al. (2016) identified this lineage of *Amyntas tokioensis* in upstate, New York at locations in Clinton and Hamilton, New York. The final Haplotype H1 had 5 individuals and was *Metaphire hilgendorfi*. Like other lineages, this lineage was identified in Japan and the South Kuril Islands (Fig. 6).

Haplotype	T1	T2	T3	A1	M1
	1.3 BT1-0.4con	1.6 BT1-40.6con	11.3 RG1-0.1	1.5 BT1-40.4con	11.7 RG1-20.1con
# of Specimens	20	8	35	11	5
Specimen	1.3 BT1-0.4con>1.4 BT1-20.14con>1.7 BT1-0.3con>17.4 BT1-10.2con>17.5 BT1-10.4con>17.6 BT1-20.3con>22.3BT10.3 MTFcon>22.6 RG1-10.5con>23.5 BT1-40.17con>25.3 BT1-0.1con>25.4 BT1-10.1con>5.3 RG1-0.4con>5.5 RG1-0.9>5.5RG0.9M13con>BTI 29.5.con>BTI 37.10con>BTI 37.3con>BTI 37.4con>BTI 37.5con>BTI 38.5.con				

Figure 5 shows the specimens in each distinct haplotype. Each haplotype is labeled in the same order found in Fig. 6, with the number of specimens as well as specimens labels below.

<i>A. tokioensis</i> 1	<i>A. tokioensis</i> 2	<i>A. tokioensis</i> 3	<i>A. agrestis</i>	<i>M. hilgendorfi</i>
Japan	Japan	Japan	Japan	Japan
—	South Kuril Islands	South Kuril Islands	South Kuril Islands	South Kuril Islands
Arkansas and Missouri	Northeast U.S	Northeast U.S	—	—

Figure 6 shows the distribution of each haplotype. Each haplotype was found in Japan, *Amyntas* species native range. Four of the haplotypes were found in the South Kuril Islands of Russia, just north of Japan. All *Amyntas tokioensis* haplotypes have been sampled in the United States, as well as locally in New York.

Discussion

Amyntas species as successful invaders of the Hamilton College campus

Our results show that all three species of *Amyntas* that are found in North American invasive jumping worm populations have successfully invaded the Hamilton College campus. These earthworms have potential to be highly invasive (Cameron & Bayne 2009; Novo et al. 2015) and the presence of all three *Amyntas* at Hamilton is most likely aided by their high mobility, reproductive rate, and parthenogenesis (Chang et al. 2021). Whether *Amyntas* species are present at other sites on the Hamilton College campus remains to be seen. Research suggests that these species invade along corridors and pathways that have high levels of human activity (Cameron & Bayne 2009; Novo et al. 2015). *Amyntas tokioensis* and *Amyntas agrestis* have been previously identified in upstate New York by Schult et al. (2016) and Chang et al. (2018). These studies did not identify *Metaphire hilgendorfi* in the Utica area, however, and this study represents the first local record of *M. hilgendorfi*.

Population structure and dynamics on the Hamilton College campus

Jumping worms were more prevalent at the Bundy Café site compared to the Root Glen site. While *Amyntas tokioensis* and *Metaphire hilgendorfi* were present at both study sites, *Amyntas agrestis* was only present at the Bundy Café site. It is possible that introduction at both sites occurred independently, through different introduction pathways. The sites are separated by 0.34 miles, a pathway, and a river, so spread directly and solely from one site to another is unlikely (Fig. 4). Thus, the presence of *A. agrestis* may indicate a more frequent introduction events, different in species composition, at the Bundy site

On the other hand, assuming that *Amyntas agrestis* was also introduced to the Root Glen site at some time in the past, competitive exclusion could account for the distinct composition of *Amyntas* species at both study sites and lack of *Amyntas agrestis* in the Root Glen. Previous *Amyntas* studies have suggested that these species compete for ecological niches leading to competitive exclusion (Chang et al. 2018). The universal prevalence and abundance of *Amyntas tokioensis*, also seen on the Hamilton College campus, suggests the species possesses qualities which make it dominant in forested landscapes (Schult et al. 2016; Chang et al. 2018; Chang et al. 2021). This species could be outcompeting other *Amyntas* species for resources. Novo et al. (2015) found a negative correlation between invading *Amyntas* species. Meaning high abundance of one *Amyntas* correlated to low abundance of another. Novo et al. (2015) mentioned this could be due to competition or preferences in soil, altitude, and habitat. Co-

invasion between the three species is possible as all have rapidly colonized North America. Invasive jumping worms can have detrimental effects to nutrient cycling and soil productivity (McCay et al. 2020; Chang et al. 2021). This could also benefit *Amyntas* species because disturbed environments are much more susceptible to invasion by non-native species (Cameron et al. 2008). *Amyntas* species have an annual life cycle, a broad diet, low predation and parasitism rates, and are epigeic allowing them to outcompete native species (Chang et al. 2016; McCay et al. 2020; Chang et al. 2021). Primary invasions by all three species could be followed by periods of resources partitioning and niche segregation.

Haplotype diversity

The haplotype diversity of *Amyntas tokioensis* as compared to the other two species indicates that there may have been multiple introduction events for this species. Or, that a single introduction event from an origin population that possessed (only) all three haplotypes. It is also possible that the same haplotype may be introduced multiple times. Low haplotype diversity suggests a low occurrence of introduction events as well as high rates of competition between lineages (Cameron et al. 2008). Multiple introduction events of the same haplotype, however, could support *Amyntas tokioensis* dominance in local communities. Individuals are regularly introduced along corridors, pathways, or in soil samples such as compost or mulch (Cameron et al. 2008; Cameron & Bayne 2009). The suitability of horticultural materials for *Amyntas* might also explain the sympatric occurrence of these lineages. Schult et al. (2016) also identified three lineages of *Amyntas tokioensis* in central New York as well. However, only one haplotype from Schult et al. (2016) was identified on the Hamilton College campus out of 24 total haplotypes.

Haplotype lineages

All five haplotypes were identified to occur in their native range of Japan (Fig. 5), and thus originated from there. The three *Amyntas tokioensis* haplotypes were also identified outside of their native range. T1 was identified in Arkansas and Missouri, T2 was identified in the northeast United States, and T3 was identified in upstate New York (Fig. 6). This suggests these species are spread widely throughout South Asia and are regularly transported internationally via anthropogenic transportation. The relative recency of studies such as Shekhovtsov et al. (2018), show how human-mediated transportation pathways, like as airplanes

or cargo barges, can rapidly accelerated the rate of dispersal of non-native species (Schult et al. 2016; McCay et al. 2020). The presence of haplotype T1, T2, and T3 in Arkansas, Missouri, and the northeast United States shows that *Amyntas tokioensis* is very well established (Fig. 6). T3, being the most abundant lineage at Hamilton, is also found throughout the United States (e.g., Schult et al. 2016), suggesting that this lineage is extremely well established as an invasive species in North America.

Introduction events

Again, *Amyntas agrestis* was first discovered in Maryland in 1939 (Chang et al. 2021). *Amyntas tokioensis* and *Metaphire hilgendorfi* were first discovered in 1947 and 1948, respectively, in New York City (Chang et al. 2021). The discovery and sampling history of *Amyntas agrestis* suggests that the species has invasively spread significantly northward, yet still is becoming established in the area of central NY. *Amyntas tokioensis* and *Metaphire hilgendorfi* have invaded northwest into upstate New York in the past decade based on sampling conducted in Chang et al. (2017) and Schult et al. (2016). *Metaphire hilgendorfi* has recently been established in the Utica area as it was not present in Chang et al. (2017). However, slight distinctions in morphological characteristics and a lack of sampling data could be contributing to the gaps in knowledge as to the extent of these species spread (McCay et al. 2020; Nouri-Aiin et al. 2021).

Conclusion

Larger importance

Since the introduction of invasive jumping earthworms to the United States almost a century ago, invasive megascolecidae have had wide-ranging effects on the biotic and abiotic environment. *Amyntas agrestis*, *Amyntas tokioensis*, and *Metaphire hilgendorfi*, have been identified as invasive species of particular interest because of their ecological impacts. As these earthworms continue to colonize the northeast, an effective understanding of their invasion and movement patterns is fundamental to environmental and conservation approaches.

Phylogeographic research is one method through which the invasion history of populations can be investigated to further create a conservation approach to address their future spread.

Collecting and identifying specimens using DNA barcoding is essential for phylogeographic analysis into invasion history and species distribution. Hopefully, phylogeographic analyses can be used to help combat the spread of *Amyntas* throughout the U.S. Locally, further sampling could be completed to gain a holistic understanding of *Amyntas* populations.

Further Research

Sampling could be completed in a wider array of environments around the Hamilton College campus to gain a greater understanding about how *Amyntas* species are introduced. A deep investigation into the origins of these five unique haplotypes could also aid in understanding how/where invasive jumping worms have been introduced to the Hamilton College campus. Because of the relatively low diversity of haplotypes, which indicate a low invasion rate, longitudinal studies could be conducted to investigate population *Amyntas* population dynamics. Studies such as Novo et al. (2015) indicate these dynamics are still mostly unknown.

References

- Bohlen, P. J., Scheu, S., Hale, C. M., McLean, M. A., Migge, S., Groffman, P. M., & Parkinson, D. (2004). Non-native invasive earthworms as agents of change in northern temperate forests. *Frontiers in Ecology and the Environment*, 2(8), 427-435.
- Cameron, E. K., Bayne, E. M., & Coltman, D. (2008). Genetic structure of invasive earthworms *Dendrobaena octaedra* in the boreal forest of Alberta: Insights into introduction mechanisms. *Molecular Ecology*, 17(5), 1189-1197.
- Cameron, E., & Bayne, E. (2009). Road age and its importance in earthworm invasion of northern boreal forests. *Journal of Applied Ecology*, 46, 28-36.
- Chang C., Snyder B., & Szlavecz, K. (2016). Asian pheretimoid earthworms in North America north of Mexico: An illustrated key to the genera *Amyntas*, *Metaphire*, *Pithemera*, and *Polypheretima* (Clitellata: Megascolecidae). *Zootaxa*, 4179(3), 495-529.
- Chang, C., Johnston, M. R., Görres, J., Dávalos, A., McHugh, D., & Szlavecz, K. (2018). Co-invasion of three Asian earthworms, *Metaphire hilgendorfi*, *Amyntas agrestis* and *Amyntas tokioensis* in the USA. *Biological Invasions*, 20(4), 843-848.
- Chang, C., Bartz, M., Brown, G., et al. (2021). The second wave of earthworm invasions in North America: biology, environmental impacts, management and control of invasive jumping worms. *Biol Invasions*.
- Chang, C. & Chen, J. (2005). Taxonomic status and intraspecific phylogeography of two sibling species of *Metaphire* (Oligochaeta: Megascolecidae) in Taiwan. *Pedobiologia*, 49.
- Corn, L.C., Buck, E.H., Rawson, J., & Fischer, E. (1999). Harmful Non-Native Species: Issues for Congress. *Congressional Research Service Issue Brief*, RL30123.
- Cristescu, M.E. (2015). Genetic reconstructions of invasion history. *Mol Ecology*, 24, 2212-2225.
- Dai, Q., Gao, Q., Wu, C., Chesters, D., Zhu, CD., & Zhang, AB. (2012). Phylogenetic reconstruction and DNA barcoding for closely related pine moth species (*Dendrolimus*) in China with multiple gene markers. *PLoS One*, 7(4).
- (DEC) Department of Environmental Conservation. Nuisance & Invasive Species: Animals, Plants, and Aquatic Life. New York State.

- Decaëns, T., Porco, D., Rougerie, R., Brown, G., & James, S. (2013). Potential of DNA barcoding for earthworm research in taxonomy and ecology. *Applied Soil Ecology*, 65, 35-42.
- Geller, C., Meyer, C., Parker, M., Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, 13(5), 851-861.
- Hajibabaei, M., Singer, G., Hebert, P., & Hickey, D. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics*, 23(4), 167-172.
- Laushman, K. M., Hotchkiss, S. C., & Herrick, B. M. (2018). Tracking an invasion: Community changes in hardwood forests following the arrival of *Amyntas agrestis* and *Amyntas tokioensis* in Wisconsin. *Biological Invasions*, 20(7), 1671-1685.
- Lustig, A., Worner, SP., Pitt, JPW., Doscher, C., Stouffer, DB., & Senay, SD. (2017). A modeling framework for the establishment and spread of invasive species in heterogeneous environments. *Ecol Evol*, 7, 8338-8348.
- McCay, T., Brown, G., Callaham, M., Chang, C., Dávalos, A., Dobson, A., Görres, J., Herrick, B., James, S., Johnston, M., McHugh, D., Minter, T., Moore, J., Nouri-Aiin, M., Novo, M., Ortiz-Pachar, J., Pinder, R., Richardson, J., Snyder, B., & Szlavecz, K. (2020). Tools for monitoring and study of peregrine pheretimoid earthworms (Megascolecidae). *Pedobiologia*, 83.
- McLean M., & Parkinson D. (2000). Field evidence of the effects of the epigenic earthworm *Dendrobaena octaedra* on the micro-fungal community in pine forest floor. *Soil Biol Biochem*, 32, 351-60.
- Novo, M., Cunha, L., Maceda-Veiga, A., Talavera, J., Hodson, M., Spurgeon, D., Bruford, M., Morgan, A., & Kille, P. (2015). Multiple introductions and environmental factors affecting the establishment of invasive species on a volcanic island. *Soil Biology and Biochemistry*, 12.
- Nouri Aiin, M., Schall, J., Keough, C., Wen., Gorres, J. (2020). Identifying the unidentifiable: A PCR multiplex protocol for the diagnosis of invasive pheretimoid earthworm species, verified by morphological and barcode identification. *Applied Soil Ecology*, 161.

- Rius, M., Turon, X., Ordóñez, V., & Pascual, M. (2012). Tracking Invasion Histories in the Sea: Facing Complex Scenarios Using Multilocus Data. *PloS one*, 7.
- Schult, N., Pittenger, K., Davalos, S., & McHugh, D. (2016). Phylogeographic analysis of invasive Asian earthworms (*Amyntas*) in the northeast United States. *Invertebrate Biology*, 135(4), 314-327.
- Shekhovtsov, S., Sundukov, Y., Blakemore, R., Gongalsky, K., Peltek, S. (2018). Identifying earthworms (*Oligochaeta*, *Megadrili*) of the Southern Kuril Islands using DNA barcodes. *Animal Biodiversity and Conservation*, 41, 9-17.
- Student, N. R. C. A., & Marchand, E. (n.d). Invasive Earthworms Impact on Red-backed Salamanders.
- Vitousek P.M. (1990). Biological Invasions and Ecosystem Processes: Towards an Integration of Population Biology and Ecosystem Studies. *Ecosystem Management*, Springer.

Appendices

Appendix 1

Reynolds Lab (ver. 2/23/22)

ISOLATION OF GENOMIC DNA FROM SAMPLES IN RNALATER OR EtOH USING QIAGEN DNEASY KIT

Day1: allow 1h for microdissection, 1h weighing & digestion set-up (4–6 specimens)

Day 2: allow 1.5h for DNA isolation

Turn on incubator and Mini Dry Bath - takes time to warm up!

Wear gloves and lab coat!

Use filter tips!

Use project-specific kit!

Project _____ Date _____ Initials _____

Reference: DNeasy Blood & Tissue Kit (50); #69504; Date & Kit lot# _____

Sample Name	Taxonomic name, and relevant info:	Size of sample?	What left after O/N digestion?

Note: Use P1000 for everything except Proteinase K (use P20).

1. Spray work area with RNase Away followed by 70% EtOH → Microdissect ~15mg of tissue, add to 1.5ml MF tube, labeled on lid with specimen name. **Avoid drying specimen!**
2. Add 180µl ATL buffer + 20 µl Proteinase K to 1.5ml tube → Add specimen → Vortex 15 sec → Tape tubes horizontally to rack and place on rocker in 56°C oven overnight.

Start time:

End time:

2

3. After O/N digestion: Vortex tubes and record appearance of digested contents → keep warm in Mini Dry Bath at 56°C while set up.
4. Use large MF tube rack (or 2 racks) to organize specimen tubes, spin columns (+ collection tube), 2 sets of new collection tubes, and 2 sets of MF tubes
 - Label spin columns on lid only with sample names only
 - Label 1st set of MF tubes on the lid with sample names only
 - Label 2nd set of MF tubes on the frosted side and lid with:

1) Specimen name	3) Date
2) "gDNA"	4) Your initials
5. Add 400µl AL buffer (*check AL buffer bottle is labeled as having been pre-mixed 1:1 with EtOH*) to ONE tube at a time → Vortex immediately to mix, and keep warm in Mini Dry Bath at 56°C.
6. Vortex all tubes for 15 sec → **Let stand @room temp (RT) x 5'** → Brief spin in mini-centrifuge
7. Set P1000 at ~700µl → Transfer digested contents to DNeasy Mini spin column + collection tube
Dispense contents/solutions onto middle of filter without touching pipette tip to filter!!
8. Spin 8000 rpm x 1' in benchtop centrifuge → Check all liquid thru; if not spin longer → Transfer column to new collection tube → Discard flow-thru and collection tube
9. Add 500µl AW1 buffer to column (*check bottle is labeled as having been pre-mixed with EtOH*) → Spin 8000 rpm x 1' → Transfer column to new collection tube → Discard flow-thru and collection tube
10. Add 500µl AW2 buffer to column (*check bottle is labeled as having been pre-mixed with EtOH*) → Spin 8000 rpm x 1' → Spin MAX speed (15,000 rpm) x 3' to dry column → Transfer column to 1.5ml MF tube [1st set] → Discard second flow-thru and second collection tube
11. Add 200µl AE buffer → **let stand @RT x 1'** → Spin 8000 rpm x 1' → Transfer column to new 1.5ml MF tube [2nd set] → Close and save 1st set of MF tubes
12. Add 200µl AE buffer → **let stand @RT x 1'** → Spin 8000 rpm x 1' → Discard spin column → Combine each 1st set elution into 2nd set tube (i.e., end with 1 MF tube for each sample) → Store gDNA in freezer (-20°C).
13. You may choose to run PCR immediately following the isolation of your DNA. Remember to still store gDNA at -20°C following PCR setup.

Appendix 2

Reynolds lab (ver. 28 Mar. 21)

PCR FROM GDNA USING HOTSTARTAQ PLUS MM KIT

Project _____ Gene _____ PCR # _____ (from PCR log)

Primer names: _____ Date _____

Thermal cycler filename _____ T_A _____ # cycles _____ Initials _____See back for location
of materials

Reaction Mix (=18µl)							
Reaction (=20µl)							
Tube #	Sample Name	R-f H ₂ O	Prim'F ¹ 25 µM	Prim'R 25 µM	HotStarTaq Plus MM ²	BSA ³ 10 µg/µl	gDNA
—	—	Alqt #	Alqt #	Alqt #	Alqt #	Alqt #	—
1	+ve ctrl	7.43µl	0.16µl	0.16µl	10µl	0.25µl	2µl gDNA
2	-ve ctrl						2µl water
3							2µl gDNA
4							
5							
6							
7							
8							

Tube #	Sample Name	Total reaction mix factor	X (# tubes + 1)
9			
10		HotStarTaq Plus MM	_____ µl
11		Primer F	_____ µl
12		Primer R	_____ µl
13		BSA	_____ µl
14		RNase-free H ₂ O	_____ µl
15		Total reaction mix vol.	_____ µl
16			

¹ Final reaction concentrations of primers = 0.2µM² Qiagen (2x): final reaction concentrations of HotStarTaq Master Mix components (at final reaction volume of 20µl):

✓ 1U DNA Polymerase (per 20µl) = 0.05 U DNA Polymerase/µl

✓ 1x Buffer (per 20µl)

✓ 200µM dNTPs (per 20µl) = 10µM dNTP/µl

✓ 1.5mM MgCl₂ (per 20µl) = 75nM MgCl₂/µl³ Final reaction concentrations of BSA = 0.125 µg/µl

Reynolds lab (ver. 28 Mar. 21)

Procedure (~1 h to start of thermocycler program):

1. Spray down surfaces with RNase Away and then 70% EtOH
2. Thaw ALL tubes
3. Mix reagents by flicking; mix gDNA by gently inverting 3–4X
4. Briefly spin all reagent tubes down w/ benchtop minicentrifuge → Place on 0.5ml-tube rack on ice
5. Set up Reaction Mix in 1.5ml tube → Note order of reagents (add RNase-free water last)
6. Flick tube to mix → Spin down w/ benchtop minicentrifuge
7. Set up reactions on ice using 0.5ml thin-wall (TW) tubes (labeled w/ PCR # & Tube #, sample name)
8. Add 18µl Reaction Mix to each labeled TW tube
9. Add 2µl of DNA or H₂O (for negative control) to each labeled TW tube
10. Briefly spin all reaction tubes down w/ benchtop minicentrifuge
11. Place in thermocycler and run appropriate program⁴
12. PCR reactions / tubes are stable in thermal cycler at end of cycles (“4°C for ever”)
13. Press PROCEED to move to next step (END) → press PROCEED again to return to home screen
14. Move PCR reaction tubes to fridge for overnight storage before gel protocol to confirm PCR product
15. Note specimen gel lanes and attach photo confirmation results below
16. If successful, send reactions for sequencing
17. Store PCR reaction tubes in -20°C freezer for long-term storage

PCR #

Lane	1	2	3	4	5	6	7	8	9	10
Tube #	Ladder	+ve	-ve							Ladder
Tube #	Ladder	+ve	-ve							Ladder

Label lanes above and tape gel image below:

Materials:

- Ice bucket, ice, cooled 0.5ml-tube rack
- 0.5ml hard-wall (HW) tubes
- 0.5ml thin-wall (TW) tubes
- p10 pipette & filter tips
- p20 pipette & filter tips
- p200 pipette & filter tips
- HotStarTaq *Plus* Master Mix (MM) (numbered 100-µl aliquots in freezer)
- Forward primer (numbered 5-µl aliquots in freezer)
- Reverse primer (numbered 5-µl aliquots in freezer)
- BSA (numbered 10-µl aliquots in fridge)
- RNase-free/nuclease-free water (numbered 0.5-ml aliquots in fridge)
- gDNA (in your labeled box in freezer) (+ve control [HAW] in primer box)
- Benchtop minicentrifuge
- Thermal cycler

⁴ Turn on thermocycler at back of unit → select “RUN” → press PROCEED → select MAIN folder → select HOTSTART program → confirm 30 L block, thin tube, 20µl volume, calculated temp control. Active timer indicates program running; hold right SELECT arrow for run time progress (HOTSTART program runs ~4:45 h; can leave overnight)

Reynolds lab (ver. 3/07/21)

TESTING PCR PRODUCTS: MAKING AND RUNNING 1% AGAROSE GELS

1. Take 250 ml flask + stir bar from left wall cabinet & drawers
 - Tare flask on balance, and ADD: **0.7 g Agarose**
 - Then ADD: **70 ml SYBR™ Safe stain (0.5X TBE soln.)**
[left wall shelves; use 100 ml graduated cylinder to transfer]
2. Microwave on one minute [to stop boiling, open door]
 - Take out and swirl at 30 sec... and continue thru 1 min **[Use orange gloves!!!]**
 - Swirl gently, then harder
 - Microwave for another minute: let boil 2–3 more times
 - Swirl after each boil; ~15 sec
 - Let cool on stir plate [speed 3–4], until can hold bottom of flask in bare hand
[~13 mins; do not over-cool!!!]
3. While gel cools, get gel cast ready:
 - Slide cast into larger, electrophoresis apparatus so that rubber of cast is touching plastic wall of apparatus. **[Should be a tight fit!!!]**
 - Select comb for # of needed wells (10) for gel and fit into top and middle slots of cast (assuming cutting gel to making 2 separate gels)
 - If 1 comb will not give you enough wells you can just use both rows of wells of one gel during one run.
4. Pour into bottom end of gel tray, corner furthest away from combs (in case of bubbles)
 - Make sure no bubbles... if there are any, touch with P200 tip to remove
 - Use kimwipe to clean up last drop on flask – pour gel residue into kimwipe
[Tip to benchtop waste bags, kimwipe to trash]
 - **CLEAN immediately:** hot water & rinse thoroughly with R/O water → Rinse stir bar in hand so doesn't go in drain
5. Let gel harden ~15 min → should be cloudy/opaque
6. Once hardened, remove combs by pulling out gently and **straight up!** **[Don't rock side to side!]**
 - remove cast from apparatus **[wiggle cast out of apparatus]** → slide gel onto saran wrap
7. Can save whole or part of gel for later:
 - Cut with razor blade to save part of gel
 - Wet a paper towel with 0.5X TBE buffer and place in Tupperware
 - Wrap up gel in saran wrap & place on wet paper towel in Tupperware
 - Label Tupperware: # of wells, initials, date, for your use later or not?
 - Store in fridge
8. Place gel in cast with wells near black lead (left side of apparatus) **[gels "ALWAYS RUN TO RED"]**
 - Pour 0.5X TBE buffer into electrophoresis apparatus until gel is covered completely.
 - Keep adding buffer until you see a smooth liquid surface at eye level.
[typically edges of wells will be highest part of gel].

Reynolds lab (ver. 3/07/21)

9. Load **5 μ l of 100 bp ladder** in *at least* first lane; if room, place ladder in first and final lane of gel [100 bp ladder stock in -4°C fridge] [black surface under gel rig helps to see wells].
10. Load **2.7 μ l PCR rxn + 1 μ l of 6X loading dye** [Loading dye is @RT by gel station]
 - Place 1- μ l dye dots on parafilm
 - Add 2.7 μ l PCR rxn
 - Mix in pipette tip (pushing to second stop) and pull up all of mix to transfer (same pipettor/tip)
 - Load into empty gel slots following PCR rxn number order – complete lane table on PCR worksheet
11. Run gel at ~100–110V) for ~20 mins; loading dye should have migrated about half way.
[Keep an eye on it!]
12. Image gel in Bio-Rad Chemidoc Xrs+ System; tape image onto back of PCR worksheet.

Reynolds lab (vers. 2/24/21)

GEL IMAGING USING BIO-RAD CHEMIDOC XRS+ SYSTEM AND IMAGE LAB SOFTWARE

1. Turn on ChemiDoc XRS+: power switch lower back left side (should see green power light come on) if not on already.
2. Turn camera on: switch on unit on benchtop to left of ChemiDoc XRS+ (should feel camera vibrating)
3. Turn on desktop computer (button center front) if not on already
4. userid "administrator" should appear, leave password box empty; press "enter"

5. Open "Image Lab" app
6. Choose New Protocol
7. In dialog box "Protocol1":
 - a. Select "Gel Imaging"
8. Under "Application" select "Nucleic Acids" and "SYBR Safe"
 - a. Should now read:
 - i. "Standard Filter (Filter1)" (can see filter 1 selected at top of gel imager; do not touch)
 - ii. "UV Trans illumination"
9. Under "Imaging Area" select "Select gel type" and "Bio-Rad Mini Ready Agarose Gel"
10. Under "Image Exposure" select "...automatically optimize..." and "Intense Bands"
11. Under Display Options select "Highlight saturated pixels" and Image Color "Gray"
12. Select "Position Gel"
 - a. Filter Position dialog box: set to Filter 1; confirm "OK"

13. Open bottom drawer of Gel Imager and position gel in center of viewing area, long axis oriented left to right, and close gel drawer. Open box above drawer to position if necessary.
14. Adjust zoom on gel for picture
15. Select "Run Protocol"
16. Insert lane lines, etc., if necessary
17. Save image to "Reynolds lab" folder
18. Print image

19. Exit program
20. Remove gel, throw in trash (assuming used SYBR Safe, not ethidium bromide), clean transilluminator surface
21. Turn off camera, leave ChemiDoc XRS+ and computer on if weeknight, turn off if weekend night

Reynolds lab (ver. 14 Feb 2022)

SENDING PCR PRODUCTS TO GENEWIZ FOR SEQUENCING

1. Login to Genewiz.com
2. Choose Sanger sequencing → For DNA type, select “PCR product, unpurified”
3. For Service Type, select “Premix” from drop-down menu if ExoSap-IT process complete; choose “Custom” if not enzymatically purified (Genewiz will do the “cleanup”)
4. For Service Priority, select “Standard”
5. For Purification Type, select “Enzymatic”
6. Provide number of samples (# templates being sent) (# reaction will be double if two primers)
7. For order name, include PCR number and date
8. For order comments, leave blank unless previously discussed
9. For Special ID, leave blank.
10. Select “Tube view” if <48 tubes; select “Plate view” if >48 → select “plate 1” → name plate (order name), vertical view
11. DNA name = sample name
12. Length = 500–1000bp
13. For Genewiz primer (Skip “My primer”), type & select M13F & M13R per cell
14. Save & review → confirm → add to cart
15. Payment by Department Biology cc → check out → print order pdf & ship with samples
16. Label tubes as instructed by Genewiz (see sheet in microtube drawer)
17. Pipette 5µl of PCR product per reaction into each tube, typically 10µl total for forward and reverse primer sequencing (one tube per template [two reactions, one per primer] if asking Genewiz to clean up with enzymatic purification; separate tubes for each reaction per template if we purify).
18. Wrap up and send by Fed Ex in small box (note Genewiz acct number if sufficient samples); collection on campus ~10am, in Clinton (near Firehouse) at 5:45pm.

Appendix 6

Reynolds lab (ver. 30 Nov 2021)

CONTIG / CONSENSUS CREATION AND SEQUENCE IDENTIFICATION USING UGENE AND BLAST

1. All result files from Genewiz are in the shared drive. You may have to download the ones you are working on to your local computer to access from UGENE; upload your work to shared drive.

Contig creation in UGENE

2. Download the freeware UGENE (<http://ugene.net/download.html>) or use on lab computer. If your Apple computer balks at downloading, control-click the dmg installer icon, then choose Open from the shortcut menu.
Read more here:
<https://support.apple.com/guide/mac-help/open-a-mac-app-from-an-unidentified-developer-mh40616/mac>
 3. For each specimen: open reverse-primer Fasta sequence file (xM13R.seq) in UGENE; note first several nucleotides to record original orientation (no way for program to remember this)
 4. Actions > Edit > Replace the whole sequence by > Complement (3'–5') sequence
 5. Actions > Edit > Replace the whole sequence by > Reverse (3'–5') sequence
 6. Select all sequence (Command-A; dialog box) > Action > Export > Export selected sequence region > File format FASTA > Click "..." box and Save as "[specimen name]rev.com.fa" > Save direct strand >
 7. Save with corresponding F&R sequences; strictly group under same root specimen-number name. (Might be better to create a folder for each specimen; you'll have 5 items for each specimen.)
-
8. In UGENE, File > Open > highlight forward and revcomp.fa sequences > Open > Join sequences into alignment (This saves a "merged_document.aln" file in your sequences folder which you can ignore and later delete.)
(You now see your two sequences with color-coded nucleotides, with little congruence.)
 9. Actions > Align with MUSCLE > Align (use default settings)
(Should be almost total congruence; if not, then check orientation of rev.com.fa sequence)
 10. Actions > Copy > Copy consensus with gaps > File > New document from text > paste consensus @ "Paste data here" > Paste specimen # @ "Sequence name" > check "Save immediately" > Click "..." to right of "Save sequence to file" > Under "Save as..." type "con" and note folder destination > Click "Save"
 11. Returning to "Create sequence dialog box > insert specimen#con to sequence name both places ("Sequence name" and "Save to file") > Click "Create"

Sequence identification in BLAST

12. Delete ends: 'N's and most gaps (or find & delete primers)
13. Copy consensus (right click; copy with gaps).
14. Go to BLAST at NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)
(It's possible to BLAST within UGENE; NCBI site is my interface preference.)
15. Choose Nucleotide BLAST; paste consensus sequence; job title = specimen lot number

Reynolds lab (ver. 30 Nov 2021)

16. Use default settings except optimize for “Somewhat similar sequences (blastn)”
17. Click “Blast”; should get new status window: “searching”; may take a minute or two before results pop up.
18. Download Description Table (.csv): genus/ species names, and info (query coverage, % identity, accession #, etc.); use specimen number as name and keep with specimen sequences.

UGENE citation:

Okonechnikov K, Golosova O, Fursov M, & the UGENE team (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28: 1166–1167. doi:10.1093/bioinformatics/bts091

BLAST citation:

Altschul SF, Gish W, Mille, W, Myers EW, & Lipman DJ (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410. doi: 10.1016/S0022-2836(05)80360-2

Some readings to help interpret BLAST results:

<https://www.ncbi.nlm.nih.gov/books/NBK1734/>

<https://medium.com/computational-biology/how-to-interpret-blast-results-ec304216fd5>

<https://support.nlm.nih.gov/knowledgebase/category/?id=CAT-01239>

<https://www.ncbi.nlm.nih.gov/books/NBK62051/>

Reynolds lab (ver. 24 Apr 2021)

CONSENSUS SEQUENCE ALIGNMENT AND NJ-TREE CONSTRUCTION USING UGENE

All sequence files from Genewiz are in the shared drive. You may have to download the ones you are working on to your local computer to access from UGENE; upload your work to shared drive.

∇ In UGENE, click on "Open" and that brings you to your folders (you may need to download the consensus seqs first, I'm not sure). Select all the consensus sequences at once, then click open.

∇ Another dialog box should open, and you should choose the "Join sequences into an alignment" option. Click "Save document" box and note the new name (modify) and location. Click "open."

∇ A new window with the color-coded sequences should appear. Note all the "N"s to the left, and the gray bars at top and bottom that indicate degree of identity across all the sequences at each site.

∇ Choose Actions> Align> Align with MUSCLE. Leave all defaults as is, and click Align. You should see the degree of identity at all sites increase dramatically, as indicated by the gray bars at the top and bottom.

∇ Use your cursor to select the columns at the left end of the sequence matrix, into the matrix as far as includes all the "Ns." Then, Actions> EDIT> Remove selection. Move the sequence navigator bar at the bottom of the window to the opposite (right) end of the matrix. Repeat removal of all columns that include an "N." This might be up to about 50 columns/ positions. If it is over a hundred, try aligning with MUSCLE again. If still a problem, contact me.

∇ Choose Actions> Tree> Build Tree, you get to the Build Phylogenetic tree dialog box. Under "Tree Building Method" choose "PHYLIP Neighbor Joining" (default)

∇ Under the "Distance Matrix" tab: for "Distance Matrix Model" choose "Kimura," check the gamma distribution box, and leave Transition/ transversion ratio at the default of 2.00

∇ Go to "Display options" tab and select "Display tree in new window." Note name of tree and location to which it is being saved. Click "Build."

∇ Click the little tree tab on the right side. Under Tree view choose "Phylogram"; UNclick "Show distances"; you can play with Branch width and height to get a nice spread of your tree.

∇ Actions> Export Tree image> Whole tree as SVG.

∇ The saved svg file can be opened in a drawing program to manipulate with colors, etc. I suggest Inkscape which is free on the web (<https://inkscape.org/>).

∇ Send me the tree. We may run a bootstrap analysis to get measures of support for the branches.

Reynolds lab (ver. 31 Mar. 22)

HAPLOTYPE ANALYSIS

A **haplotype** (**haploid genotype**) is a group of **alleles** in an **organism** that are inherited together from a single parent.^{[1][2]}

Many organisms contain genetic material (**DNA**) which is inherited from two parents. Normally these organisms have their DNA organized in two sets of pairwise similar **chromosomes**. The offspring gets one chromosome in each pair from each parent. A set of pairs of chromosomes is called **diploid** and a set of only one half of each pair is called haploid. The haploid genotype (haplotype) is a genotype that considers the singular chromosomes rather than the pairs of chromosomes. It can be all the chromosomes from one of the parents or a minor part of a chromosome, for example a sequence of 9000 **base pairs**.

However, there are other uses of this term. First, it is used to mean a collection of specific alleles (that is, specific **DNA** sequences) in a cluster of tightly linked genes on a **chromosome** that are likely to be inherited together—that is, they are likely to be **conserved as a sequence** that survives the descent of many generations of reproduction.^{[3][4]} A second use is to mean a set of **linked single-nucleotide polymorphism** (SNP) alleles that tend to always occur together (i.e., that are **associated statistically**). It is thought that identifying these statistical associations and a few alleles of a specific haplotype sequence can facilitate identifying *all other such* polymorphic sites that are nearby on the chromosome. Such information is critical for investigating the genetics of common **diseases**, which in fact have been investigated in humans by the **International HapMap Project**.^{[5][6]} Thirdly, many human genetic testing companies use the term in a third way: to refer to an individual collection of specific mutations within a given genetic segment; (see **short tandem repeat** mutation).

<https://en.wikipedia.org/wiki/Haplotype>

Our model: **Schult et al. 2016**

- Unique haplotypes (also polymorphic sites?)
 - FASTA Tools Unique Sequences program
- Haplotype diversity (also nucleotide diversity?)
 - Arlequin program
- Haplotype relationships
 - Arlequin program (minimum spanning tree)
 - HapStar program (visualized)
- Uncorrected pairwise p-distances among lineages
 - MEGA program
- Within-lineage pairwise distances
 - Arlequin program

Look up and review each:

- unique haplotypes
- polymorphic sites
- haplotype diversity
- nucleotide diversity
- haplotype relationships
- minimum spanning tree
- uncorrected pairwise p-distances among lineages
- within-lineage pairwise distances

Reynolds lab (ver. 31 Mar. 22)

Unique haplotypes

FASTA Tools Unique Sequences program

https://www.ncbi.nlm.nih.gov/CBBresearch/Spouge/html_ncbi/html/fasta/uniqueseq.cgi

1. Create trimmed alignment (MUSCLE) of consensus sequences in UGENE
2. Save subalignment in FASTA format
3. Confirm FASTA format in TextEdit program. See FASTA format here:
https://www.ncbi.nlm.nih.gov/CBBresearch/Spouge/html_ncbi/html/barcode_file_check/help.html
4. Submit to FASTA Tools Unique Sequences program at site above
5. Read output FASTA file; sequences are sorted into unique haplotypes; sequence name listed for each unique haplotype.

Useful vocabulary:

A barcode FASTA define has only one '>' (at its start), followed by an identifier unique to the sample (its UID) For example:

>Genus_species_ID

Polymorphic sites

I believe the MEGA program will give this to us – later.

Haplotype diversity

We will have this number from the FASTA Tools Unique Sequences program

Nucleotide diversity

The PopART program will give us this.

Haplotype relationships

Schult et al. (2016) used the Arlequin program (minimum spanning tree) and HapStar program (visualized).

Instead we are using PopART:

<http://popart.otago.ac.nz/index.shtml>

1. Download appropriate for computer
2. Using UGENE, save alignment in nexus format
3. Confirm nexus format here: <http://popart.otago.ac.nz/examplenex.shtml>
4. Create traits block for location (we do not need geotags block or trees block)
 - o Bundy, Root
 - o If present= 1, if absent= 0
 - o so if sequence found at Bundy= 1,0; if sequence found at Root= 0,1
 - o we will try this way and see what it tells us; may try different trait formats
5. Open in PopART
6. View Traits, Alignment
7. View Network; choose Minimum Spanning Network
8. Play with visualization; we want to get an indication of % location for each haplotype, and the # of haplotypes at each location.

Uncorrected pairwise p-distances among lineages

Schult et al. (2016) used the MEGA program and we will also, later in week...

Within-lineage pairwise distances

Schult et al. (2016) used the Arlequin program; I believe MEGA will give this to us also.