**Author:** Matthew E. Aiello-Lammens

**Title:** Reconstructing the historical spread of *Frangula alnus* using herbarium records

**Affiliation:**

**Email of corresponding author:** matt.lammens@gmail.com; 631-327-2371

**To-do:**

\* Slim intro material to that appropriate for Biological Invasions

\* Read through the paper to find the main gist (see below)

\* Update refs – see tag “fral-herbariam” in mendeley

\* Re-examine hypotheses – possibly needs better framing

**What is the main gist of this paper?**

As the paper stands now, I think it’s very much a descriptive analysis of the invasion of buckthorn. That’s not bad *per se*, but I don’t think it’s terribly interesting.

What I would like to refocus the paper on is how the duration of a lag phase differs based on spatial extent of the study region. This will require some re-analysis, but I think would be a worthwhile contribution.

**Abstract**

Analysis of herbaria records can aid in understanding the processes that govern non-native species invasions, allowing for an examination of patterns of the spread in a novel range.  I used herbaria records to investigate the rate of spread and pattern of establishment for the invasive shrub *Frangula alnus* (Glossy Buckthorn) in northeastern and middle North America.  *F. alnus* is a perennial woody species of concern to land managers throughout the invaded range. Accession records were collected from online databases of herbaria throughout North America and from direct requests to herbaria curators, resulting in >700 records of *F. alnus* covering a temporal range from ca. 1880-Present and a spatial range broadly covering the entire invaded area in northeast North America. I addressed unequal sampling effort by comparing temporal and spatial patterns of *F. alnus* accessions to patterns in a group of ecological similar native species. Current understanding of the earliest stages of the *F. alnus* invasion is that the plant was likely first introduced in southern Ontario (near London), however my examination finds that initial introductions appear to have been greatly separated geographically, ranging from southern Ontario to coastal New York and New Jersey.  Such large spatial separations are likely the result of multiple introductions. Trends in record collection in time and space show the rate of spread of *F. alnus* was initially slow, then increased rapidly during the early 20th century, and reached a relatively constant rate of spread in the later 20th century. Examining the spread of this species at the continental scale, there is little evidence that it experienced an extended lag phase between phases of establishment and rapid spatial spread, in contradiction to previous reports. Rather, it appears that *F. alnus* has steadily increased in area of occupancy since ca. 1920 to the present.

**Introduction**

A remaining gap in our understanding of species invasions is a full understanding of population dynamics during the transition from establishment of self-sustaining populations in a novel region to the rapid growth in abundance and expansion of area of occupancy characteristic of invasive species. The time from establishment to rapid spread is generally noted as being longer than a generation time for many invasive species. This period is thought to be one in which population growth, both in numbers and area, are relatively small, and is commonly referred to as the lag phase (Kowarik 1995, Crooks and Soulé 1999, Sakai et al. 2001, Crooks 2005, Pyšek and Hulme 2005, Theoharides and Dukes 2007). Because the lag phase is most often associated with a period early in the invasion process, management actions taken during this time have the potential to be most affective at mitigating the negative impacts of a species invasion.

To understand processes governing population dynamics during a lag phase, it is first necessary to identify the lag phase. Quantifying the lag phase for a species has many challenges. Assuming a non-native species establishes in an environment suitable for growth and reproduction, and ignoring potential positive density dependence effects (i.e. Allee effects), population size should initially increase exponentially ~~(Figure 1A)~~. Exponential growth is described as *N(t) = N(0) Rt*, where *t =*time-step, *N(t)* = the population size at time-step *t*, and *R* is the population growth rate. A visual examination of population size through time for an exponentially growing population, suggests that there is period of time early in the population growth trajectory during which the population size remains relatively similar to *N(0)* ~~(Figure 1A; *R* = 1.2 for the black dots)~~, and this period is followed by a transition to one in which the population size rapidly increases. Crooks and Soulé (1999) describe this as “the shallow portion early in the growth curve when the population is growing relatively slowly in absolute number” and define it as the **inherent lag**. This pattern is common to all exponential growth curves, and thus all growing populations should appear to have at least an inherent lag phase. However, because a mathematical definition of inherent lag is lacking, it is challenging to apply this concept. In the case of exponential population growth, a plot of the *log* of population size versus time shows a linear relationship (Figure 1B; *R* = 1.2 for black dots). The population growth rate is constant through time and there is no distinct transition point that can define the shift from the inherent lag phase to the population explosion phase.

While an inherent lag cannot be explicitly defined mathematically, it serves as a contrast to a population growth curve that shows an **extended lag**. An extended lag phase is one in which population growth early in the growth curve is less than it is later. The grey dots in Figures 1A and B are growth curves for a population whose initial growth rate is *R* = 1 and increases to *R =* 1.2 during the first 20 time steps. As is the case for population growth with constant *R*, the growth curve of population size versus time is non-linear (Figure 1A). However, in this case the growth curve of the log of population size versus time is also non-linear while *R* is increasing. The non-linear portion of the curve, where the slope is more shallow and the curve is concave up, is considered evidence for an extended lag (Crooks and Soulé 1999). Patterns of population growth for many invasive species show evidence of an extended lag phase. The potential factors causing this are not well understood; several ecological and evolutionary processes may be involved (Mack et al. 2000, Sakai et al. 2001, Pyšek and Hulme 2005, Gurevitch et al. 2011). An extended lag may be explained by time constraints intrinsic to population growth and establishment, such as generation time and time to first reproduction (i.e., the time required for a population to achieve a stable age distribution). However, many observed extended lag phases are longer than can be explained by these processes. In a recent study of weedy species in New Zealand, Aikio et al. (2010a) point out that while there are several proposed mechanisms that may explain extended lags, relatively little empirical work has been done to examine their validity. Among the most well studied mechanisms in a theoretical context are the role of evolutionary adaptation of the invader during the lag phase, waiting time until a disturbance event avails resources to the invader, and the role of complex dispersal mechanisms (for a general review see (Crooks 2005). Historical biodiversity data from herbaria and museums could be applied to address this lack of empirical support and further our understanding of the population dynamics of the lag phase. These data can be used to re-construct patterns of spatial spread of invasive species, which are linked to increases in population size.

Retrospective spatial analyses can yield insights into the ecological processes involved in the spread of non-native invasive species in novel regions. A rich data source for these types of analyses on plant species is the specimen holdings of herbaria. Information from herbarium records are used in several studies to estimate species rates of spread through time and space (e.g., (Delisle et al. 2003, Salo 2005, Miller et al. 2009, Crawford and Hoagland 2009, Newbold 2010, Aikio et al. 2010a, 2010b, Lavoie 2012). These data have also been used to investigate native species range expansions beyond historical limits due to changing ecological conditions (e.g., Feeley et al. 2011; Feeley 2012). Despite their utility, analyzing these data presents a number of challenges; and determining whether a species had an extended lag phase is not a trivial task. One substantial challenge is that herbarium records may have been collected with unequal sampling effort in time and/ or space, resulting in biases in datasets compiled from these records. For example, historically there are documented periods of high and low specimen collection for herbaria in general (Prather et al. 2004). There may also be herbarium specific trends, such as a peak in collection activity following the opening of a herbarium or an emphasis on regional specimen collection. Other biases may emerge because of issues of convenience. A noted pattern in natural history collections is the large number of specimens collected near museums, herbaria, botanical gardens, and academic centers, as well as urban areas in general, where there is a higher concentration of naturalists (Hijmans et al. 2000, Reddy and Dávalos 2003, Kadmon et al. 2004). For plants collection in particular, this is sometimes referred to as the “botanist effect”. These potential biases make it difficult to determine if observed trends in herbarium collections (e.g., increased number of records through time) are indicative of changes in the population size or range size of the species of interest, or rather, representative of trends in overall specimen collection. For example, Catling and Porebski (1994) found that observations of the plant *Frangula alnus* were generally concentrated around urban areas from the time this species was first observed in southern Ontario, late 1800s, up to the 1970s. This pattern may be the result of the fact that botanists were located near these areas, and thus their collections tended to come from these areas, or it may be that *F. alnus* grows well in ecological conditions near urban areas (e.g., disturbed environments). It is important to distinguish these patterns to use these data for understanding ecological processes of species invasions.

Another challenge in using herbarium records is that the data are generally more conducive to studying spatial spread, rather than population increase. These processes are related, but it should not be assumed that the rate of spatial spread is equivalent to the rate of population growth. This is problematic because whereas exponential growth is widely accepted as a null model for local population growth, there are many plausible null models for areal growth (i.e., spatial spread) for range expanding species. One commonly used null model is a reaction-diffusion model, which results in a description of spatial growth as a function of the radius of the population. The area of occupancy for a species spreading randomly on a landscape from a central point should increase geometrically, with an exponent of 2. With this assumption, the square root of the area of occupancy through time for a spreading plant should have a linear relationship with respect to time, analogous to linearity after log transform for exponential growth. While other null models have been proposed, the square root transformation has been shown to be applicable to most of these and used to examine historical data on spatial spread (Crooks and Soulé 1999).

Given the promise of the utility of natural history collections in general (Graham et al. 2004, Anderson 2012), and herbaria in particular (Lavoie 2012), it is not surprising that the effects of unequal sampling effort have been discussed in recent studies using herbarium records. Most methods used to address these effects compare the trends in the distribution of records for a species of interest to other species that have similar habitat requirements – i.e. associated species (Miller et al. 2009). For example, Catling and Porebski (1994) compared the pattern of collection records of *F. alnus* to that of *Rhamnus alnifolia* and showed that botanists were in fact collecting specimens well outside of urban areas during the time frame of the introduction and early spread of *F. alnus*. They thus concluded that *F. alnus* was in fact primarily located near urban areas during this time. More recent developments in analysis methods make it possible to account for potential biases in a more robust manner than a simple visual comparison (e.g. (Delisle et al. 2003, Aikio et al. 2010a). These methods have been used to reconstruct patterns of range expansion for both non-native and native plants (e.g., Miller et al. 2009; Larkin 2011). At least one of these methods (i.e., Aikio et al. 2010a) was specifically developed and applied to identify the existence, and estimate the duration, of lag phases for invasive plants. Adopting and modifying these methods, I address some of the unanswered questions regarding the introduction and spread of the non-native invasive plant *F. alnus* (Glossy buckthorn)*.*

*Frangula alnus* is purported to have had an extended lag phase (Catling and Porebski 1994, Frappier et al. 2003b, Larkin 2011), and various mechanisms have been proposed to explain this observation. Howell and Blackwell (1977) suggested that the rapid expansion of *F. alnus* into Ohio, following an extended period of presence in the eastern United States may be associated with the rapid expansion of the non-native European starling. Lending support to this idea, European starlings have recently been linked to the spread the non-native invasive plant *Celastrus orbiculatus* (Oriental bittersweet) (Merow et al. 2011), which is also a woody fruit-bearing plant. However, Catling and Porebski (1994) pointed out that the spread of *F. alnus*  in southern Ontario preceded the presence of European starlings, suggesting other mechanisms must be at play. While Howell and Blackwell’s speculation concerns the spread of *F. alnus* throughout the state of Ohio (>110,000 km2), Frappier and colleagues (2003b) speculate on a mechanism causing an observed extended lag phase in the invasion of a 250 m2 forest plot, suggesting that the lag may be due to “early selection and adaptation” to the local ecological conditions. Despite these speculations, it remains unclear as to whether *F. alnus* did in fact have an extended lag phase.

In this study, I examined the range expansion of *F. alnus* throughout its novel range of northeastern North American. I compiled a dataset of historical occurrence observations of *F. alnus* consisting primarily of herbarium records, but also including observations noted in the scientific literature. Using these data I calculated metrics related to the rate of spatial expansion of *F. alnus* throughout northeast North America employing modified methods of (Delisle et al. 2003 and Aikio et al. 2010a) to account for unequal sampling effort of herbarium records. I hypothesized that my findings would support previous claims that *F. alnus* experienced an extended lag lasting from the time of its introduction (ca. 1860) to the early 1970s. The findings and patterns established in this chapter provide an important foundation for further analyses carried out in *Chapter 4*, in which I use linked population and species distribution models to examine the population dynamics of the *F. alnus* invasion. Many previous studies have used herbarium records to estimate rates of spread and examine spatial patterns of spread, but they have largely been limited to a regional focus. In this study I investigated the range expansion of *F. alnus* throughout the entire novel range. I expanded upon the methods of Delisle et al. (2003) to allow for incorporation of data from multiple herbaria, while still accounting for unequal effort in sampling. Access to herbaria records and other historical biodiversity is increasing as more collections are digitized and made available to the public. Having techniques to combine information from desperate sources, so as I present here, will help scientists use these data to investigate important standing questions.

**Methods**

*Collection of presence records*

I assembled a dataset of occurrence records for *F. alnus* to investigate its spread throughout northeast North America. I collected historical presence records by 1) searching publicly available on-line databases of several herbaria, 2) requesting *F. alnus* record information from curators and collection managers of various herbaria, 3) searching the Global Biodiversity Information Facility (GBIF), and 4) carrying out a literature search for all North American localities, using the keyword search terms “Frangula alnus” or “Rhammus frangula” in ISI Web of Science database. I only included presence records with information on the year and location (county level or finer) of observation in my dataset. I accessed herbarium records from 30 herbaria via publically available on-line databases (Table 1). Many herbaria manage on-line databases that include records from affiliated, but separate, herbaria. For example, the Wisconsin State Herbarium is located at the University of Wisconsin at Madison. A search of the database for this herbarium yields accession records from this herbarium (WIS), as well as the University of Wisconsin at Stevens Point herbarium (UWSP), the University of Wisconsin at Green Bay herbarium (UWGB), and the Morton Herbarium (MOR). Therefore, while 30 herbaria are represented in my dataset, a smaller number of databases were used to acquire these records. I contacted curators and collections managers at several herbaria directly to inquire about whether there were records for *F. alnus* in their collections. Curators or collection managers at three herbaria, Rutgers Chrysler Herbarium, Miami University of Ohio Herbarium, and Carnegie Mellon Herbarium, provided me with the record information for all *F. alnus* or *R. frangula* in their collections. In addition to collecting records from herbarium databases, I collected records accessed searching the Global Biodiversity Information Facility (GBIF), the Canada Biodiversity Information Facility (CBIF), and the Invasive Plant Atlas of New England (IPANE). Records collected from all three of these biodiversity databases were carefully examined to exclude records accessed in other searches.

I searched for historical records of *F. alnus* presence reported in published journal articles using the ISI Web of Science database. I used keyword search terms of “Frangula alnus” and “Rhamnus frangula”. From this literature search, I included 5 presence records reported by Howell and Blackwell (1977), Taft and Solecki (1990), and Catling and Porebski (1994). Howell and Blackwell (1977) investigated the history of the spread of *F. alnus* into and throughout Ohio, and reported that the first recorded observation was from Lake County, Ohio in 1927 (confirmed via search of Ohio State University Herbarium). Taft and Solecki (1990) reported that *F. alnus* was first recorded in the state of Illinois in 1912 (Sherff 1912, as cited by Taft and Solecki 1990) in Cook County (confirmed via search of Field Museum Herbarium). Catling and Porebski (1994) investigated the spread of *F. alnus* into and throughout southern Ontario. I used data reported in this paper to identify time and location of the three earliest records of *F. alnus* presence in this region.

*Georeferencing of historical presence records*

Most of the presence records in the compiled dataset did not have latitude and longitude values for where the specimen was collected. However, many records had descriptive information, which I used to assign latitude and longitude values to the collection location. This process is commonly called georeferencing. I primarily used the GoogleEarth software (Google Inc.) to georeference records (*sensu* Garcia-Milagros & Funk 2010), however some cases required additional Internet searches on named locations, such as unique building names or geographic features. The descriptive location information allowed for varying degrees of precision in the final latitude and longitude value assigned to a record. At a minimum, all records of specimens collected in the United States had enough information to be assigned to the county in which the specimen was collected. For records with *only* county level information, I assigned the US Census Bureau designated latitude and longitude values to the record. For records of specimens collected in Canada, I relied on the information included with the specimen to assign county equivalent values. To achieve finer spatial resolution, some information allowed for assigning values for the township the specimen was collected in, whereas others allowed for assigning precise locations. I assigned location uncertainty values using guidelines from the Biogeomancer Consortium (BioGeomancer Consortium 2006) to all of the georeferenced records. I carried out analyses using different subsets of the compiled dataset based on the spatial resolution of the presence record (described further below).

*Number of records through time*

Trends in the number of records added to an herbarium through time were used to infer rates of spread for non-native invasive species and to determine if a species experienced an extended lag phase (*sensu* (Pyšek and Prach 1993, Aikio et al. 2010a). I calculated the cumulative number of records reported per year in the dataset of compiled historical *F. alnus* presences and plotted the log of the cumulative number of records through time against years. I fit linear, quadratic, and cubic regression lines to these data, and used a likelihood ratio test to determine the best-fit model. I calculated an annual rate of growth for the cumulative number of presence records by dividing the cumulative number of records at year *t +* 1 by the cumulative number of records at year *t*. The mean rate of growth was calculated as the geometric mean of the annual growth rates. Additionally, I calculated 10-year moving window geometric mean growth rates, which minimizes the influence of extreme inter-annual fluctuations in growth rates. This yielded more consistent rates of growth for the earliest period of the invasion, during which time calculations are based on a relatively small number of records.

*Area of occupancy through time*

I examined the spatial pattern and rate of range expansion of *F. alnus* throughout its novel range using the dataset of compiled historical *F. alnus* observations described above. To do this I created a map of equal sized grid cells for the area of interest (Longitude: -97.0 – -62.0 degrees; Latitude: 38.0 – 48.0 degrees) using the Quantum GIS software (QGIS Development Team 2012). Each grid cell was 5 x 5 arc minutes, which is generally reported as 10 x 10 km. Because a unit of longitude is smaller at higher latitudes than at lower latitudes, the area of each grid cell decreases towards the poles. The area for each grid cell ranges from ca. 67 km2 to ca. 57 km2. Because *F. alnus* occurrences were widely distributed throughout the study region, I assumed the effects of these differences in grid cell area would be negligible on average. I used the R statistical programming environment (R Core Team 2012) with additionally functions from the “raster”, “rgdal”, and “dismo” packages (Hijmans 2012, Keitt et al. 2012, Hijmans et al. 2013) respectively) to carryout all other analyses. Each record in the compiled dataset was assigned membership to one grid cell based on its latitude and longitude value. If a grid cell contained at least one occurrence of *F. alnus*, it was considered occupied. I calculated multiple occupancy measures, including total area occupied per decade and the cumulative area occupied from time of first introduction to the present. In calculating the latter measure, I assumed that once a grid cell was classified as occupied, it would not later be classified as unoccupied. Similar measures have been used in other studies investigating the spread of invasive species (e.g., (Pyšek and Prach 1995, Weber 1998, Delisle et al. 2003). Further, given the difficulty of removing *F. alnus* and the lack of reported successful eradications, I believe this is a reasonable assumption. I calculated the rate of growth for the area of occurrence analogously to how I calculated the rate of growth of the number of records (see *Number of records through time*). I substituted the cumulative number of records with the cumulative number of occupied grid cells and plotted the square root of the cumulative number of grid cells versus time (years). Assuming areal growth is a random diffusion process, this relationship should be linear (\*\*CITATION NEEDED\*\*). A deviation from linearity that is concave up indicates a period of time during which spatial spread is slower than random diffusion. Likewise, a concave down curve indicates a period of during which spatial spread is more rapid than random diffusion.

*Occupied counties through time*

Many records contained only enough spatial information to georeference the collection location to the county in which it occurred. Furthermore, georeferencing records to county requires substantially less time and effort than higher levels of precision. Thus, analyzing spatial patterns of herbarium records at the county level makes the compilation and use of large datasets more achievable given limited time and resources, while still providing insights into the patterns and processes of species invasions (e.g., Barney 2006). Similar to the calculations of *Area of occupancy through time*, I calculated the cumulative number of counties occupied through time for the compiled dataset. Again, I assumed that once *F. alnus* was found in a county, the county was henceforth considered occupied. I calculated the growth rate for the cumulative number of counties occupied in similar manner to how I calculated the rate of growth of the number of records (see *Number of records through time*).

*Accounting for unequal sampling effort in time and space*

The potential effects of unequal sampling effort complicate the interpretation of observed patterns in historical presence locations. One method for addressing these effects is to compare trends in records of a species of interest to trends for a species, or group of species, whose range and population size should be in equilibrium with their environment (e.g., native species) (Delisle et al. 2003). I chose the following group of associated native species, each of which has similar habitat requirements as *F. alnus*: Speckled Alder (*Alnus incana*), Smooth Alder (*Alnus serrulata*), Alderleaf Buckthorn (*Rhamnus alnifolia*), Meadow Willow (*Salix peiolarisi*)*,* Witch Hazel (*Hamamelis virginica* (syn. *macrophylla*)), and White Ash (*Fraxinus Americana*). These species represent woody plants that are likely to be observed in ecological conditions where *F. alnus* is observed. As noted above, Catling and Porebski (1994) compared the distribution of record collections for *R. alnifolia* to that of *F. alnus* as a comparison. Similarly, *Salix peiolaris* was used in a paired comparison with *F. alnus* in a study on the effects of invasive and native species on wetland species diversity (Houlahan and Findlay 2004). The other species in this group are found in ecological conditions conducive to the growth of *F. alnus* (personal observations; Little et al. 1980; Sibley 2009). To construct the associated species dataset, I searched GBIF for all records that were located within the area of interest described in *Area of occupancy through time*. Additionally, I collected all records for these species reported in the following herbaria databases: University of Wisconsin, Ohio State University, University of Minnesota, the Morton Arboretum Herbarium, Michigan State University, and Brooklyn Botanic Gardens. These records were georeferenced to the county level. Grouping records for all of these species into one dataset, I calculated the metrics for the number of records through time, the area of occupancy through time, and the number of counties occupied through time, as described above.

I compared the trends for *F. alnus* and the group of associated species for both area of occupancy through time and counties occupied through time in a similar manner to Delisle et al. (2003), who used herbarium records to investigate the spread of six non-native species in Quebec, Canada. They compared trends in area of occupancy through time by dividing the cumulative number of occupied 10x10 km grid cells for each non-native plant by the cumulative number of cells occupied by a group of five native species, yielding a proportion of non-native to native occupied cells for each year of their study period. It is important to note that a major assumption in the interpretation of this analysis is that spatial growth of native plants in the herbaria dataset does not represent the spread for these plants *per se*, but rather represents the increase in spatial coverage of herbaria records in general (i.e., an increase in the number of locations where samples are collected). Thus, if the proportion of occupied cells between non-native and native plants is increasing through time, this represents periods during which the spatial coverage of records for the non-native plant outpaces the background increase in spatial coverage. This can be interpreted as the result of the non-native plant rapidly spreading in space.

In this study, I was interested in examining spatial spread in a larger region than that studied in (Delisle et al. 2003), as well as in using records from multiple herbaria across the full naturalized range of *F. alnus*. This lead to several presence records for the group of associated species being located in areas where *F. alnus* has not been observed, most likely due to the larger niche breadth of some of the associated species compared to *F. alnus,* (e.g., *Fraxinus americana* (White Ash), as determined by examining the Biota of North America Program – North American Plant Atlas; Kartesz 2013). Add citation into final draft: Kartesz, J.T., The Biota of North America Program (BONAP). 2013. *North American Plant Atlas.* (http://www.bonap.org/napa.html). Chapel Hill, N.C. [maps generated from Kartesz, J.T. 2013. Floristic Synthesis of North America, Version 1.0. Biota of North America Program (BONAP). (in press)]. Including these records in my analysis had the effect of increasing the area of occupancy measures for the group of associated species compared to the possible area of occupancy for *F. alnus*. Similarly, the historical presence record dataset I manually constructed included records of *F. alnus* from institutions from which associated species records were not collected (e.g., Miami University Herbarium and Rutgers Herbarium). The records acquired from these institutions were largely geographically restricted (for example, primarily located in the states of Ohio and New Jersey for the institutions mentioned), and these locations were not always represented well in the larger group of associated species dataset. Therefore, including these records has the effect of increasing the area of occupancy of *F. alnus* compared to the possible area of occupancy for the group of associated species. To account for both of these issues, I examined the ratio of cumulative area of occupancy of *F. alnus* to the cumulative area of occupancy of the associated group of species, limiting the records used to an area of coarse spatial overlap for both datasets. I defined the spatial overlap by creating a map of equal sized grid cells, again for the area of interest defined above, where each grid cell was 30 x 30 arc minutes (i.e., 0.5° or approximately 50 x 50 km). As described in *Area of occupancy through time*, each record was assigned membership to one 30’ grid cell based on its latitude and longitude value. I then constructed restricted *F. alnus* and associated species occurrence datasets, in which only records that occurred in a 30’ grid cell occupied by at least one record from *both* datasets during the study period. Using these restricted datasets, I calculated the ratio of the increase in the cumulative area of occupancy of *F. alnus* to the group of associated species. In a separate analysis I compared the cumulative number of counties occupied through time, while accounting for similar concerns regarding falsely sampling regions in space that are unsuitable to *F. alnus*. I only included records from counties that were occupied at some time by both *F. alnus* and one of the associated plants. In this case, the ratio of the cumulative number of counties occupied at the end of the study period had to equal 1.0. The growth rates for the cumulative number of grid cells occupied and the cumulative number of counties occupied were compared between *F. alnus* and the group of associated species. To compare the growth rates I divided the annual growth rate of *F. alnus* records by the annual growth rate of the entire group of associated species.

*Comparison of number of records through time*

In addition to comparing trends in the areal increase of *F. alnus* and the associated species, I also compared trends in the cumulative number of records through time. For each year of the study period, I divided the cumulative number of *F. alnus* records by the cumulative number of associated species records. This correction method is similar to that proposed by (Delisle et al. 2003) and has been used in several studies to account for bias due to unequal sampling effort (e.g., Mihulka & Pyšek 2001; Aikio et al. 2010a; Larkin 2011). Additionally, to compare rates at which *F. alnus* and associated species occurrences were collected, I calculated the ratio of growth rates using both the annual and 10-year moving window average estimates. A ratio value less than 1.0 indicated that the rate at which *F. alnus* occurrences were being collected was lower than that of the associated species, while a ratio value of greater than 1.0 indicated that it was greater that that of the associated species.

**Results**

I compiled a dataset of 749 occurrence records for *F. alnus* and 5548 occurrence records for the combined group of associated species. *Frangula alnus* specimens have been collected in much of northeast and middle North America (Figure 2), and collection locations for the group of associated species was largely inclusive of where *F. alnus* was collected. The earliest dated occurrence record for *F. alnus* was an herbarium specimen collected in 1879 in Hudson County, New Jersey (accessed from CHRB). The earliest dated occurrence record for an associated species was a herbarium record for *Hamamelis virginiana* collected in 1836 in Richland County, Ohio (accessed from CM). Only 36 associated species records pre-dated the first *F. alnus* record, representing less than 1% of the associated records. A total of 14 *F. alnus* specimens were deposited in 4 separate herbaria prior to 1900. Of these records, 12 were georeferenced to the metropolitan New York City area (9 in the state of New Jersey) and the remaining 2 were collected in southern Ontario (as reported in Catling & Porebski 1994). Based on recorded specimen collection dates, the number of herbarium specimens collected for both *F. alnus* and the group of associated species has increased through time (Figure 3). For the group of associated species, there was a substantial increase in the number of records collected at the beginning of the 20th century, followed by another increase after 1950. There was a steady increase in the number of records collected per decade from the time of the first recorded presence of *F. alnus* in the United States to the end of the 20th century. For both groups, there was a dramatic decline in the number of records at the start of 21st century, which is a pattern that has been observed elsewhere (Prather et al. 2004).

The log cumulative number of records through time indicates that the rate at which records were added to herbaria for both *F. alnus* and the group of associated species does not fit an exponential growth relationship (i.e., a linear relationship) (Figure 4A). The best-fit regression models of log cumulative number of records versus time (years) were a cubic polynomial regression for *F. alnus* (R2 = 0.989, df = 124, P << 0.05) and a quadratic polynomial regression for the group of associated species (R2 =0.979, df = 174, P<<0.05 respectively), both departing from a linear relationship. The cumulative number of records for the group of associated species increased steadily from the late 1800s to the present. The cumulative number of records for *F. alnus* increased rapidly from the late 1800s to approximately 1930. Ratios of the annual growth rates and ten year moving average (geometric mean) growth rates show that the rate at which *F. alnus* collections occurred was generally slower than that of the group of associated species prior to approximately 1920, but faster during most of the 20th century (Figure 4B). The ratio of the cumulative increase in records of *F. alnus* to the associated species corroborates this pattern, indicating that the increase of *F. alnus* records was greater than the associated species from 1920 to the present (Figure 4C).

The cumulative number of occupied 5 arc min grid cells for *F. alnus* and the associated species, constrained to overlap within 30 arc min grid cells, increased through time for both datasets (Figure 5). A plot of the square root of cumulative occupied grid cells versus time supports a departure from a simple diffusion model of spatial spread in historical presence records (Figure 5A). If a random diffusion process governed the spatial spread of herbaria record locations, a linear regression would be the best-fit model of the square root of cumulative records versus time (Crooks and Soulé 1999). In contrast, based on likelihood ratio tests, the best-fit regression models of the square root of the cumulative number of grid cells occupied, with year as the predictor variable, were a quadratic polynomial regression for *F. alnus* (R2 = 0.996, df = 129, P << 0.05) and a cubic polynomial regression for the group of associated species (R2 = 0.990, df = 140, P << 0.05), both indicating a departure from linearity. The rate of increase in occurrence records was low for both *F. alnus* and the group of associated species until 1890, after which the number of occupied cells rapidly increases for the group of associated species. In contrast, the number of occupied cells continued to increase slowly for *F. alnus* until approximately 1920, as is demonstrated by the calculated ratio between the two growth rates (Figure 5B). Prior to 1915, with the exception of the earliest years of the *F. alnus* invasion, the rate at which new grid cells were considered occupied by ­*F. alnus* was lower than the corresponding rate for the group of associated species. From 1915 to the present, the rate at which grid cells are considered occupied by *F. alnus* was consistently higher than the corresponding rate of growth for the associated species. The ratio of the cumulative number of occupied grid cells of *F. alnus* to the associated species shows that after approximately 1910 the rate at which *F. alnus* increased in area of occurrence was greater than that of the group of associated species (Figure 5C).

Trends in the cumulative number of counties occupied were similar to the results of the analysis of the cumulative number of occupied grid cells. As noted in *Occupied counties through time*, the compiled records dataset for both *F. alnus* and the associated species was subset to include records from counties that were occupied by both during the study period. The number of counties where herbarium samples were collected increased very rapidly during the late 19th to early 20th century, as indicated by rapid increase in the cumulative occupied counties for the associated species. The number of counties where *F. alnus* was found increased very slowly during the early part of the 20th century, but rapidly after 1940 (Figure 6A). Assuming counties were selected as locations for herbarium specimen collections at random, the relationship between the square root of the cumulative number of occupied counties and time (years) should be linear for the associated species. However, this is not the relationship for either the associated species or *F. alnus* (Figure 6A). Compared to linear and quadratic regression models, a cubic polynomial regression model was the best fit model for both the associated species (R2 = 0.967, df = 166, P < 0.001) and *F. alnus* (R2 = 0.997, df = 130, P < 0.001), as determined by a likelihood ratio test. The difference in the rate of growth of cumulative occupied counties between *F. alnus*­ and the associated species shows a similar pattern to that reported for cumulative occupied grid cells (Figure 6B). The rate at which counties are considered occupied is slower for *F. alnus* than the associated species early in the invasion history (prior to 1900) and faster during most of the 20th century (Figure 6B). The ratio of occupied counties of *F. alnus* to those occupied by the associated species indicates that early in the invasion history, *F. alnus* was less frequently collected in newly sampled counties, but that for most of the 20th century the number of counties occupied by *F. alnus* has increased more rapidly than the number of counties occupied by associated species (Figure 6C). *F. alnus* specimens were collected on average 48 years after at least one of the associated species had been collected in a county. ~~While the distribution of years between sampling of an associated species and *F. alnus* is quite wide (standard deviation = 44; Figure 7), the first year an associated species was collected in a county was significantly earlier than the first year *F. alnus* was sampled in that county (paired t-test, t = -15.57, df = 196, P << 0.05).~~

\*\*\*\* I am not convinced that Fig 7 is important. The only piece of useful and relevant information that can be gleaned from it is that for most counties, some sampling occurred many years prior to finding Buckthorn there. If this stays in the paper at all, it could potentially go into a supporting information, but more likely, I can simply state the overall result, probably in the discussion.

**Discussion**

*Early observations and likely region of first introduction*

The question of *how* *F. alnus* was introduced remains unanswered. Based on its long history of use as a medicinal plant (United States Pharmacopeial Convention 1910) and evidence that it was planted as an ornamental shrub (Sherff 1912) Add reference to Sudworth and Fernow: (Sudworth, G. B. and B. E. Fernow.  1891.  *Trees of Washington, D.C.*, compliments of the Forestry Division. Geo. B. Sudworth, botanist ; B. E. Fernow, chief.  Washington, D.C.: Bell Lithographing Co.  [18] pp.; 2 folding plans; 14 x 22 cm.), it is likely the *F. alnus* was purposely planted in many locations in North America. In fact, low fertility cultivars of *F. alnus* are still available for purchase (Jacquart and Knight 2010). A thorough investigation of seed catalogs and nursery records from the late 19th century may shed more light on when and where it was planted, but was beyond the scope of this project. As is the case with many non-native species, there is little documented evidence of the time(s) or place(s) that *F. alnus* was introduced. The most frequently referenced time of first introduction is the mid-19th century (Catling and Porebski 1994, Possessky et al. 2000, Frappier et al. 2003a), however at least one source suggests it occurred prior to 1800 (Converse 1984). This source may mistakenly be referring to the closely related *Rhamnus cathartica* (Common Buckthorn; Kurylo et al. 2007). The location of first introduction has been reported variably as “the eastern states” (Howell and Blackwell 1977) and Canada (Frappier et al. 2003a). My findings suggest that a potential location of first introduction was the metropolitan New York City area and areas of New Jersey along New York Harbor. Specimens for 12 of 14 records dating from before 1900 were collected in this region. *F. alnus* was present in this region for more than a decade before it was first reported in southern Ontario, countering the claim that southern Canada was the location of first introduction. However, locations in southern Ontario account for the remaining 2 of 14 records dating before 1900, indicating that introduction into the region was likely early during the spread of *F. alnus*. It is interesting to note that the region the first recorded observations come from was a large shipping and port area, raising the possibility that shipping played a role in an accidental introduction. Many non-native plants have been introduced to port areas by the unloading of solid ballast, with seeds mixed in with rocks and other materials used as ballast (Sorrie 2005, Barney 2006). However, this region historical had a high population density, thus there also may have been more people who purposely planted *F. alnus*.

*Accounting for unequal sampling effort in historical biodiversity collections*

Using historical biodiversity collections to reconstruct patterns of species presence presents many challenges. One of the most common is the presence of unequal sampling effort (Reddy and Dávalos 2003, Graham et al. 2004, Kadmon et al. 2004, Boakes et al. 2010). For herbaria in particular, there are many reasons for unequal sampling effort in specimen collection, some of which have been discussed above (i.e. the ‘botanist effect’). Another potential cause is herbaria specific emphasis on regional collections. For example, the Oberlin Herbarium collection (housed in the Ohio State Herbarium) includes several thousand records primarily collected from within the state of Ohio. Thus, confronted with numerous records of *F. alnus* in Ohio in the compiled dataset, it is hard to discern if conditions in Ohio are favorable for establishment and spread of buckthorn, or if this region is simply better sampled than others. There may also be unequal sampling effort associated with taxonomic grouping. There are numerous examples of collections of orchids, bryophytes, mosses, and ferns, which are all groups of particular interest to plant collectors.

Calculating the ratio of the cumulative number of records in space and time of non-native to native plants offers a way to account for unequal sampling effort, making it possible to distinguish periods of relatively slow versus rapid spread (Delisle et al. 2003). The primary assumption in this correction method is that native species are in equilibrium with their environment prior to the collection of any records. That is, these plants have spread to all of the ecological conditions within the study range where they can survive and reproduce, and have a stable range distribution. There are some caveats to this assumption to consider, perhaps the most important being that the ecological conditions in northeast North America have not been stable over the last 130 years. There have been substantial changes in land use, resulting in changes to plant communities (Wright and Fridley 2010), as well as affecting plant invasion dynamics (Mosher et al. 2009). However, given that the species chosen for the group of associated species have similar ecological requirements as *F. alnus*, I expect that any such changes would affect trends in these species and *F. alnus* equally.

Provided the equilibrium assumption is true for the group of associated species in this study, then the rate of the cumulative number of records, grid cells, or counties occupied is not representative of the spread of these plants, but rather of the effort of herbarium specimen collectors. As such, if the ratio of cumulative records of *F. alnus* to the associated species is increasing in time, this indicates a period during which *F. alnus* is increasing more rapidly than background sampling, and thus experiencing positive growth rates. If the ratio is constant, *F. alnus* may be increasing, however it is not distinguishable from sampling effort. If the ratio is decreasing, *F. alnus* may still be increasing, but more slowly than the rate of sampling effort. Because the *cumulative* number of records was used in all three temporal trends calculated in this study, the absolute rate of change in samples cannot show a decline. Time periods that have either decreasing or stable ratio values, which precede periods of increasing ratio values, may be considered lags.

*Cumulative records through time*

All three of the calculated ratio values suggest that *F. alnus* increased since ca. 1920 even when increased sampling effort was taken into account (Figures 4C, 5C, and 6C). The ratio of sample growth rates also supports this claim. For all three trends, the ratio of ten-year average growth rates was greater than 1.0 after 1920, indicating that the number of *F. alnus* samples increased more rapidly than those of the group of associated species (Figures 4B, 5B, and 6B). Prior to 1920, both the ratio values of cumulative number of records and growth rates fluctuated greatly. This may be the result of the relatively small number of cumulative records for both *F. alnus* and the associated species during this time. The addition of a small number of records to either dataset could drastically change the ratio between them. The accumulation of *F. alnus* occurrences had a particularly large effect early in its invasion history, as is indicated by the ratio of growth rates, in which generally *F. alnus* records increased more slowly than the associated species from 1879 to the 1920s, but there are anomalous years in which the growth rate of *F. alnus* was much larger than that of the associated species (Figures 4B, 5B, and 6B). For example, when the cumulative number of records for *F. alnus* increased from 4 to 10 from 1893 to 1894, the resulting growth rate was R = 2.5. Comparatively, during this time the number of records for the associated species increased from 171 to 191, resulting in a growth rate of R = 1.12. Thus, it is difficult to determine if *F. alnus* did in fact start its invasion with a very high growth rate, then immediately slowed, or rather this result is an artifact of calculating growth rates with small sample sizes. Based on the relative consistency of the total number of records collected in each decade, there is more support for the latter interpretation (Figure 3).

*Spatial spread and area of occupancy through time*

Given the equilibrium assumption of native species, if the cumulative number of grid cells (and counties) occupied by the associated species were appropriately represented by a random diffusion process (i.e., linear when regressing square root of the cumulative number of grid cells onto time), this would imply that specimen collectors moved randomly outward from a central point, e.g. an herbarium, collecting new records as they went. Neither the cumulative number of grid cells nor counties occupied fit such a relationship (Figures 5A and 6A). In both cases, the trend can be described as concave up initially, followed by a linear trend with a steep slope, followed by a concave down curve, followed by another period of linearity. This suggests that the rate of collections increased rapidly in the mid 19th century, was high throughout the second half of the 19th century, and then decelerated during the 20th century. The cubic regression fits do not necessarily support the presumption of an early concave up portion of the curve, but do support the interpretation of a rapid rate of increase early in the collection history, followed by a deceleration in collection rate. However, more complex curve fitting such as GAM or piecewise linear regression may more closely fit these data. Regardless, it is clear that specimen collection rapidly spread across the landscape during the 19th century. Combined with the increase in the cumulative number of records during earlier part of the study period (Figures 3 and 4A), these patterns suggest that this was a period of high sampling effort, coinciding with the beginning of an intense effort to collect specimens for herbaria (Prather et al. 2004).

The earliest occurrence records of *F. alnus* were collected during this period of high sampling effort. However, it was uncommon across the study region at this time, and appears to have remained uncommon until at least the 1920s. Assuming that cumulative occurrence curves for the associated species represent the spatial spread of collection effort, and that a collector would collect *F. alnus* if it were present during a survey, it appears that the rate of spatial spread during the early 19th century of *F. alnus* was slow. This is most clearly exemplified by the trends in cumulative number of counties occupied through time. On average *F. alnus* was first observed 48 years after at least one of the associated species was observed in a county occupied by both by the end of the study period. By 1900, records for at least one of the associated species was collected in 42% (73 of 172) of counties, where as *F. alnus* was collected in 2% (4 of 172) of counties. This strongly suggests that *F. alnus* was not common at this time. However, the ratio of occurrence records shows a consistent rate of spread throughout the novel range during the following 100 years.

*Evidence for an extended lag phase*

Compared to most previous studies, here I examined the spread of an invasive species over a spatial area incorporating nearly all of the known naturalized range. Examining the relative rates of increase in occupied grid cells and occupied counties at this spatial scale, there is no compelling evidence for an extended lag phase persisting beyond the early 20th century. All three analyses suggest that *F. alnus* increased at a rate greater than sampling effort from at least 1920 on. Given that the rate of increase in record number and spatial occurrence for *F. alnus* was less than that of the associated species from 1880 to 1920, it is plausible that this period of time represents an extended lag phase. However, trends calculated for this period are based on a small number of *F. alnus* records, as is evident in the high variability in both the ratio of growth rates and the ratio of records or occupied area.

Three previous studies examined aspects of the range expansion of *Frangula alnus* using herbarium records (Howell and Blackwell 1977, Catling and Porebski 1994, Larkin 2011). Each focused on only part of the invaded range. Howell & Blackwell (1977) examined the spread of *F. alnus* (using the synonym *Rhamnus frangula*) into, and throughout, Ohio. They found evidence that it likely entered Ohio in the 1920’s, being observed first in Lake County (northeastern Ohio). The authors speculate that the spread of *F. alnus* westward throughout Ohio was facilitated by the range-expansion of the non-native European Starling. Many bird species are noted to eat the fruit of *F. alnus* and defecate undigested seeds, and European Starlings in particular have been known to eat these fruit (Howell & Blackwell 1977). The authors did not account for potential unequal sampling effort in herbarium records, but rather simply reported when and where *F. alnus* first appeared in Ohio. Catling & Porebski (1994) examined the historical spread of *F. alnus* in southern Ontario, Canada (also using the synonym *R. frangula*) and found that it was first recorded in London, Ontario in 1898 and Ottawa in 1899. Their data suggested that *F. alnus* spread to other urban centers, but primarily remained confined to these areas until the 1970’s. From the 1970’s to the early 1990’s *F. alnus* spread into natural areas outside of urban areas at an increased rate compared to the previous 70 years. The authors suggest that this observation is indicative of a lag phase in the spread of *F. alnus*, though no quantitative analysis was carried out. To address potential biases resulting from unequal sampling efforts, specifically for records collected prior to 1930, the authors visually compared the spatial distribution of the *F. alnus* records with that of native *Rhamnus alnifolia*, a species with similar habitat requirements. More recently, Larkin (2011) examined the lengths of lag phases for multiple (>200) species of non-native invasive species in the Wisconsin and the southern Lake Michigan region, applying the methods of (Aikio et al. 2010a). The earliest record of *F. alnus* was collected in 1908 in the southern Lake Michigan region, and based on quantitative analysis had a 31-year lag phase. Comparatively, it was found in southern and northern Wisconsin in 1927 and 1941, and had 36- and 15-year lag phases, respectively. These years of first regional introduction suggest that *F. alnus* spread north from the southern Lake Michigan region into more northern parts of Wisconsin. The length of the lag phase calculated for southern Lake Michigan and southern Wisconsin generally agrees with my observations for the whole range.

*Caveats and potential sources of bias*

I made several assumptions in carrying out the analyses presented here. One assumption of note is that I treated the cumulative number of occupied grid cells, and counties, as representative of the area of extent. This may not be valid if *F. alnus* invades an area, but later goes locally extinct. This is an unlikely occurrence. With one notable exception (Cunard and Lee 2008), there are no documented examples of the local extinction of *F. alnus*, either by natural processes or management actions.This supports the notion that once a location is occupied, it remains so. However, I am confident that some locations of historical *F. alnus* occurrences no longer have the species present because of changes in land use (e.g., development of once-woodland plots). This is the case for several of the records observed in the metropolitan New York region, the site of many of the earliest records. Disturbances, or removal of population from the landscape, due to anthropogenic influences may have a substantial effect on the spread of *F. alnus* throughout its novel region. Nevertheless, it can be successful in very small, isolated, and disturbed plots within a urban or suburban land-use matrix (personal observation; (Del Tredici 2010). Ultimately, the finest spatial resolution used in this study was 5 x 5 arc minutes (approximately 10 x 10 km) for the occupied grid cells through time. The spatial resolution for the number of counties through time varies, but the mean county area for all counties in the study region is approximately 1500 km2, making it more course than the occupied grid cells through time. For both resolutions, a complete local extinction of *F. alnus* is unlikely.

A challenge in using historical biodiversity collections is that most herbaria do not have electronic databases of their holdings that are easily accessed by the public (i.e., a web-base search interface) and many do not have a complete electronic database of records within the institution (Lavoie 2012). Therefore, I know there are several herbaria records that were not included in my compiled dataset because I did not acquire them. Nevertheless, given the extent of the data compilation in space and across institutions, I am confident that the patterns and trends reported are generally representative of the spread of *F. alnus*. I was able to collect records from some herbaria by directly contacting curators and collection managers. Many collection managers were happy to provide me with record information for *F. alnus*, in part because there were generally a small number of records, and some were able to provide me with information from records of the group of associated species. However, because of the large request entailed in collecting records for the associated species, I was not able to collect these records from some herbaria. This is particularly problematic because I focused my direct contact efforts on herbaria that I either knew contained *F. alnus* (e.g. Miami University Herbarium) or herbaria in regions that I thought were lacking in records collected by other means. These collection issues result in underestimates of the amount of area occupied by *F. alnus*, and further collections could potentially increase the accuracy of my results.

**Conclusions**

Despite concerns regarding the extent to which the compiled dataset approximates a complete representation of the area occupied by *F. alnus* through time, my results clearly support a rapid rate of spread for this species throughout its invasion history. Calculating the ratio of the cumulative number of grid cells and counties occupied by *F. alnus* to those occupied by the group of associated species, I was able identify time-periods associated with the expansion of *F. alnus* while accounting for potential unequal spatial and temporal sampling bias in occurrence record collection. Based on these calculations, *F. alnus* has expanded rapidly throughout its invaded range since the mid to late 1920s. Patterns of spatial spread and estimates of lag phase duration likely vary between regions. However, a quantitative analysis from one region yielded a length for an extended lag phase consistent with what was observed throughout the entire invaded range (Larkin 2011). The patterns and processes of range expansion, particularly of lag phase dynamics, likely vary depending on scale and local ecological conditions (Theoharides and Dukes 2007). Further analysis could entail using the datasets I have compiled, but restricting the calculations carried out here to regional levels. I examine these patterns from a different perspective in *Chapter 4*, where I use integrated species distribution and demographic models to investigate the local and regional population processes that result in the patterns I found in this chapter. This integration allows for an examination of how local population processes, such as individual plant survival, fruit production, and seed dispersal, propagate to regional patterns such as those discussed here.

**Table 1.** Historical presence record sources and counts. Abbreviations for “Accession Method” column: CBIF = Canadian Biodiversity Information Facility, GBIF = Global Biodiversity Information Facility, IPANE = Invasive Plant Atlas of New England, and vPlants = Virtual Herbarium of the Chicago Region.

|  |  |  |  |
| --- | --- | --- | --- |
| **Source** | **Herbarium Code** | **Record Count** | **Accession Method** |
| Harvard University Arnold Arboretum | A | 2 | IPANE |
| Acadia University Herbarium | ACAD | 8 | GBIF |
| Botanischer Garten und Botanisches Museum Berlin-Dahlem | B | 1 | GBIF |
| Brooklyn Botanical Garden | BKL | 35 | Institute website |
| Chicago Botanic Garden | CHIC | 3 | Institute website / vPlants |
| Rutgers University Chrysler Herbarium | CHRB | 10 | Provided by curator |
| Carnegie Museum of Natural History | CM | 68 | Provided by curator |
| University of Connecticut Torrey Herbarium | CONN | 85 | IPANE / GBIF |
| Field Museum of Natural History | F | 28 | Institute website / vPlants |
| Harvard University Herbaria | GH | 5 | IPANE |
| Royal Botanical Gardens | HAM | 7 | CBIF |
| University of Kansas | KANU | 3 | GBIF |
| Forest Products Laboratory | MAD | 1 | Institute website (U. of Wis.) |
| University of Massachusetts | MASS | 22 | IPANE |
| University of Minnesota | MIN | 48 | Institute website |
| Missouri Botanical Garden | MO | 2 | Institute website |
| Morton Arboretum | MOR | 43 | Institute website |
| Michigan State University | MSC | 25 | Institute website |
| Université de Montréal | MT | 22 | GBIF |
| Miami University | MU | 57 | Provided by curator |
| Yale University Connecticut Botanical Society | NCBS | 3 | IPANE |
| New England Botanical Club | NEBC | 63 | IPANE |
| New York Botanical Garden | NY | 5 | Institute website |
| Ohio State University | OS | 81 | Institute website |
| Queen’s University | QK | 2 | Institute website |
| University of Wyoming Rocky Mountain Herbarium | RM | 2 | Institute website |
| Smithsonian Institution | US | 3 | Institute website |
| University of Wisconsin – Green Bay | UWGB | 16 | Institute website (U. of Wis.) |
| University of Wisconsin – Stevens Point | UWSP | 6 | Institute website (U. of Wis.) |
| Naturhistorisches Museum Wien | W | 2 |  |
| University of Wisconsin – Madison | WIS | 85 | Institute website (U. of Wis.) |
| Yale University Peabody Museum of Natural History | YU | 4 | IPANE / GBIF |
| Literature Search |  | 5 | ISI Web of Science |

Figure 1. Population size versus time relationship for an exponentially growing population. Black dots represent constant population growth rate *R* = 1.2. Grey dots represent an increasing growth rate for the first 20 time steps from *R* = 1.0 to *R* = 1.2, then constant *R* = 1.2 from time points 20 to 40. (A) Population size versus time. Inset plot is Population size versus time for *R* = 1.2 for the first 20 time steps. (B) Log (Population size) versus time. For constant *R*, note the non-linear relationship in (A) versus the linear relationship in (B). This relationship is non-linear in both (A) and (B) for a population with an increasing growth rate (grey dots).



Figure 2. Geographic locations of collected records for *F. alnus* and a group of associated species. Red points represent records from the compiled historical presence records for *F. alnus* and green points represent records from the compiled historical presence records for the group of associated species. Some locations were assigned geographic locations based on the latitude and longitude values of counties as defined by the US Census Bureau. Red outlined box delineates the study region.



Figure 3. Total number of records collected in each decade for *F. alnus* (black bars) and the combined group of associated species (grey bars).



Figure 4. (A) Log cumulative number of records through time. Linear (solid) and cubic polynomial (dot-dash) regression predictions are plotted over the cumulative increase curves. (B) Ratio of growth rates of cumulative number of records of *F. alnus* versus associated species calculated annually (black circles) and by10 year moving window average (geometric mean) (red dots). Note that some extreme data points are not shown (those >1.2 or <0.8, but contribute to the moving window average values. (C) Ratio of the cumulative number of records of *F. alnus* to those of associated species.

**B**

**A**



**C**

Figure 5. (A) Square root of the cumulative number of grid cells through time. Shown here are the linear and polynomial regression lines for a models using year as a predictor variable and the square root of the cumulative number of grid cells occupied as the response variable. Linear regression predictions are shown for both *F. alnus* and the group of associated species. The best-fit polynomial regression fit is shown for each set (quadratic for *F. alnus* and cubic for the group of associated species) (B) Ratio of growth rates of cumulative occupied grid cells calculated annually (black points) and by 10 year moving window average (geometric mean) (red points). Outlier data not shown (>1.2 or <0.8), but do contribute to the moving window average values. (C) Ratio of square root of the 5 arc min grid cells occupied by *F. alnus* and associated species at a given time step. Occupied 5 arc min grid cells were constrained to be within 30 arc min grid cells occupied by both *F. alnus* and associated species at least once during the study period.

**B**

**A**

**C**

****

Figure 6. (A) Square root of the cumulative number of counties occupied through time. Shown here are the linear and polynomial regression lines for a models using year as a predictor variable and the square root of the cumulative number of counties occupied as the response variable. Linear regression predictions are shown for both *F. alnus* and the group of associated species (solid red and black lines respectively). The best-fit polynomial regression fit is shown for each set (cubic for both *F. alnus* and the group of associated species) (B) Ratio of the rate of growth for cumulative occupied counties calculated annually (black points) and by 10 year moving window average (geometric mean) (red points). Outlier data not shown (those >1.2 or <0.8, but contribute to the moving window average values. (C) Square root of the ratio of the cumulative number of counties occupied by *F. alnus* to those occupied by the group of associated species.

**C**

**B**

**A**

****

**\*\*\*\***Remove this figure.\*\*\*\*

**Figure 7.** Frequency of the differences in the number of years between an observation of an associated species in a county and the observation of *F. alnus* in that county.



**References**

Aikio, S., R. P. Duncan, and P. E. Hulme. 2010a. Lag-phases in alien plant invasions: separating the facts from the artefacts. Oikos 119:370–378.

Aikio, S., R. P. Duncan, and P. E. Hulme. 2010b. Herbarium records identify the role of long-distance spread in the spatial distribution of alien plants in New Zealand. Journal of Biogeography 37:1740–1751.

Anderson, R. P. 2012. Harnessing the world’s biodiversity data: promise and peril in ecological niche modeling of species distributions. Annals of the New York Academy of Sciences 1260:66–80.

Barney, J. N. 2006. North American history of two invasive plant species: phytogeographic distribution, dispersal vectors, and multiple introductions. Biological Invasions 8:703–717.

BioGeomancer Consortium. 2006. Guide to Best Practices for Georeferencing. (A. D. Chapman and J. Wieczorek, Eds.). Global Biodiversity Information Facility, Copenhagen.

Boakes, E. H., P. J. K. McGowan, R. A. Fuller, D. Chang-qing, N. E. Clark, K. O’Connor, and G. M. Mace. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biology 8:e1000385.

Catling, P. M., and Z. S. Porebski. 1994. The history of invasion and current status of glossy buckthorn, *Rhamnus frangula*, in southern Ontario. Canadian field-naturalist 108:305–310.

Converse, C. K. 1984. Element stewardship abstract for *Rhamnus cathartica*, *Rhamnus frangula* (syn. *Frangula alnus*). The Nature Conservancy, Arlington, The Nature Conservancy.

Crawford, P. H. C., and B. W. Hoagland. 2009. Can herbarium records be used to map alien species invasion and native species expansion over the past 100 years? Journal of Biogeography 36:651–661.

Crooks, J. 2005. Lag times and exotic species: the ecology and management of biological invasions in slow-motion. Ecoscience 12:316–329.

Crooks, J. A., and M. E. Soulé. 1999. Lag times in population explosions of invasive species: causes and implications. Pages 103–125 *in* O. T. Sandlund, P. J. Schei, and A. Viken, editors. Invasive species and biodiversity management. Kluwer Academic Dordrecht, The Netherlands.

Cunard, C., and T. D. Lee. 2008. Is patience a virtue? Succession, light, and the death of invasive glossy buckthorn (*Frangula alnus*). Biological Invasions 11:577–586.

Delisle, F., C. Lavoie, M. Jean, and D. Lachance. 2003. Reconstructing the spread of invasive plants: taking into account biases associated with herbarium specimens. Journal of Biogeography 30:1033–1042.

Feeley, K. J. 2012. Distributional migrations, expansions, and contractions of tropical plant species as revealed in dated herbarium records. Global Change Biology 18:1335–1341.

Feeley, K. J., M. R. Silman, M. B. Bush, W. Farfan, K. G. Cabrera, Y. Malhi, P. Meir, N. S. Revilla, M. N. R. Quisiyupanqui, and S. Saatchi. 2011. Upslope migration of Andean trees. Journal of Biogeography 38:783–791.

Frappier, B., R. T. Eckert, and T. D. Lee. 2003a. Potential impacts of the invasive exotic shrub Rhamnus frangula L. (glossy buckthorn) on forests of southern New Hampshire. Northeastern Naturalist 10:277–296.

Frappier, B., T. D. Lee, K. F. Olson, and R. T. Eckert. 2003b. Small-scale invasion pattern, spread rate, and lag-phase behavior of *Rhamnus frangula* L. Forest Ecology and Management 186:1–6.

Garcia-Milagros, E., and V. A. Funk. 2010. Improving the use of information from museum specimens: using Google Earth© to georeference Guiana Shield specimens in the US National Herbarium. Frontiers of Biogeography 2:71–77.

Graham, C. H., S. Ferrier, F. Huettman, C. Moritz, and A. T. Peterson. 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution 19:497–503.

Gurevitch, J., G. A. Fox, G. M. Wardle, Inderjit, and D. Taub. 2011. Emergent insights from the synthesis of conceptual frameworks for biological invasions. Ecology Letters 14:407–418.

Hijmans, R. J. 2012. Introduction to the’raster’package (version 2.0-08). R.

Hijmans, R. J., K. A. Garrett, Z. Huaman, D. P. Zhang, M. Schreuder, and M. Bonierbale. 2000. Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. Conservation Biology 14:1755–1765.

Hijmans, R. J., S. J. Phillips, J. R. Leathwick, and J. Elith. 2013. dismo: Species distribution modeling. R.

Houlahan, J. E., and С. S. Findlay. 2004. Effect of invasive plant species on temperate wetland paint diversity. Conservation Biology 18:1132–1138.

Howell, J. A., and W. H. J. Blackwell. 1977. The history of *Rhamnus frangula* (glossy buckthorn) in the Ohio flora. Castanea 42:111–115.

Jacquart, E. M., and T. M. Knight. 2010. Are there noninvasive cultivars of buckthorn. Indiana Nursery and Landscape News 70:16–17.

Kadmon, R., O. Farber, and A. Danin. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications 14:401–413.

Keitt, T. H., R. Bivand, and B. Rowlingson. 2012. rgdal. R.

Kowarik, I. 1995. Time lags in biological invasions with regard to success and failure of alien species. *in* P. Pyšek, K. Prach, M. Rejmánek, and M. Wade, editors. Plant invasions: General aspects and special problems. SPB Adademic Publishing, Amsterdam.

Kurylo, J. S., K. S. Knight, J. R. Stewart, and A. G. Endress. 2007. *Rhamnus cathartica*: Native and naturalized distribution and habitat preferences. The Journal of the Torrey Botanical Society 134:420–430.

Larkin, D. J. 2011. Lengths and correlates of lag phases in upper-Midwest plant invasions. Biological Invasions 14:827–838.

Lavoie, C. 2012. Biological collections in an ever changing world: herbaria as tools for biogeographical and environmental studies. Perspectives in Plant Ecology, Evolution and Systematics 15:68–76.

Little, E. L., S. Bullaty, and A. Lomeo. 1980. National Audubon Society Field Guide to North American Trees: Eastern Region. Alfred A. Knopf.

Mack, R. N., D. Simberloff, W. Mark Lonsdale, H. Evans, M. Clout, and F. A. Bazzaz. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. Ecological Applications 10:689–710.

Merow, C., N. LaFleur, J. A. Silander Jr, A. M. Wilson, and M. Rubega. 2011. Developing dynamic mechanistic species distribution models: predicting bird-mediated spread of invasive plants across northeastern North America. The American Naturalist 178:30–43.

Mihulka, S., and P. Pyšek. 2001. Invasion history of Oenothera congeners in Europe: a comparative study of spreading rates in the last 200 years. Journal of Biogeography 28:597–609.

Miller, R. J., A. D. Carroll, T. P. Wilson, and J. Shaw. 2009. Spatiotemporal analysis of three common wetland invasive plant species using herbarium specimens and geographic information systems. Castanea 74:133–145.

Mosher, E. S., J. A. Silander Jr, and A. M. Latimer. 2009. The role of land-use history in major invasions by woody plant species in the northeastern North American landscape. Biological Invasions 11:2317–2328.

Newbold, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Progress in Physical Geography 34:3–22.

Possessky, S. L., C. E. Williams, and W. J. Moriarity. 2000. Glossy buckthorn, *Rhamnus frangula* L.: A threat to riparian plant communities of the northern Allegheny Plateau (USA). Natural Areas Journal 20:290–292.

Prather, L. A., O. Alvarez-Fuentes, M. H. Hayfield, and C. J. Ferguson. 2004. The decline of plant collecting in the United States: a threat to the infrastructure of biodiversity studies. Systematic Botany 29:15–28.

Pyšek, P., and P. E. Hulme. 2005. Spatio-temporal dynamics of plant invasions: linking pattern to process. Ecoscience 12:302–315.

Pyšek, P., and K. Prach. 1993. Plant invasions and the role of riparian habitats: a comparison of four species alien to central Europe. Journal of Biogeography 20:413–420.

Pyšek, P., and K. Prach. 1995. Invasion dynamics of *Impatiens glandulifera* - a century of spreading reconstructed. Biological Conservation 74:41–48.

QGIS Development Team. 2012. QGIS Geographic Information System.

R Core Team. 2012. R: A language and environment for statistical computing. Vienna, Austria.

Reddy, S., and L. Dávalos. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. Journal of Biogeography 30:1719–1727.

Sakai, A. K., F. W. Allendorf, J. S. Holt, D. M. Lodge, J. Molofsky, K. A. With, S. Baughman, R. J. Cabin, J. E. Cohen, N. C. Ellstrand, D. E. McCauley, P. O’Neil, I. M. Parker, J. N. Thompson, and S. G. Weller. 2001. The population biology of invasive species. Annual Review of Ecology and Systematics 32:305–332.

Salo, L. F. 2005. Red brome (*Bromus rubens* subsp. *madritensis*) in North America: possible modes for early introductions, subsequent spread. Biological Invasions 7:165–180.

Sherff, E. E. 1912. Range extenstions of *Rhamnus frangula* and *Sporobolus asperifolius*. Rhodora 14:227–229.

Sibley, D. 2009. The Sibley guide to trees. Alfred A. Knopf, New York.

Sorrie, B. A. 2005. Alien vascular plants in Massachusetts. Rhodora 107:284–329.

Taft, J. B., and M. K. Solecki. 1990. Vascular flora of the wetland and prairie communities of Gavin Bog and Prairie Nature Preserve, Lake County, Illinois. Rhodora 92:142–165.

Theoharides, K. A., and J. S. Dukes. 2007. Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. New Phytologist 176:256–273.

Del Tredici, P. 2010. Wild Urban Plants of the Northeast: A Field Guide. Comstock Publishing Associates, Ithica.

United States Pharmacopeial Convention. 1910. The Pharmacopeia of the United States of America. Ninth Dece. P. Blakiston’s Son & Company, Philadelphia.

Weber, E. 1998. The dynamics of plant invasions: a case study of three exotic goldenrod species (*Solidago* L.) in Europe. Journal of Biogeography 25:147–154.

Wright, J. P., and J. D. Fridley. 2010. Biogeographic synthesis of secondary succession rates in eastern North America. Journal of Biogeography 37:1584–1596.