1 **Implementing a new approach to effective conservation of genetic diversity, with ash**
2 **(*Fraxinus excelsior*) in the UK as a case study**

3
4 **ABSTRACT**

5 Gene conservation programs help safeguard species and tangibly benefit ecological restoration,
6 agriculture, forestry, and horticulture. Here we describe a new method for deciding which and
7 how many populations and individuals to conserve ex situ, and we demonstrate the method by
8 evaluating collections of European Ash (*Fraxinus excelsior*) for an ongoing seed-banking
9 project, the UK National Tree Seed Project (NTSP). The method uses simulations and
10 geographic distribution data, but does not require genetic data. We estimate that NTSP
11 collections have captured >90% of global and locally common alleles. We identified optimal
12 sampling solutions at large and small spatial scales, and for northern isolated vs. southern core
13 populations. We also quantified genetic “points of diminishing returns” with a more precise
14 method than previous studies. This analysis revealed that (for European ash, for a goal of
15 capturing one copy of each allele) an optimal “stopping point” is approximately 35 populations,
16 10 to 30 trees per population, and 30 seeds per tree. Overall, we conclude that the NTSP
17 protocol of random sampling of at least 15 trees per population from two populations per seed
18 zone is effective. We demonstrated how collectors can adjust the number of populations,
19 individuals and seeds sampled using the concept of “genetic equivalence”, allowing projects to
20 accommodate practical or ecological constraints. Lastly we showed that for a conservation goal
21 of 50 allele copies rather than one copy, a much larger sampling effort is needed (>150
22 populations). This new approach can be tailored to any species. It is applicable to any seed
23 collection seeking to capture genetic diversity, as well as in situ gene conservation approaches.
24 We emphasize that the ability to quantitatively estimate the outcome of gene conservation
25 activities can help design, justify, or evaluate future programs.

26 1 INTRODUCTION

27 A challenge for biodiversity conservation practitioners is to be both effective and efficient- to
28 devise and implement plans that have a high probability of achieving defined outcomes within
29 the constraints of limited resources. Computational modeling, informed by real data and
30 knowledge of natural history, can help predict whether a conservation action is likely to be
31 successful and help choose among possible options. For example, for several decades,
32 population viability analysis has been performed using modeling approaches to help plan
33 optimal conservation actions like supplementation (Boyce 1992, Hoban 2014), while
34 conservation planning has been used to design optimal reserve networks (Carwardine et al.
35 2009). However, methodologies for ensuring effective conservation of a species' evolutionary
36 potential (i.e. genetic diversity, the variability and distribution of genes) are relatively
37 undeveloped in spite of the known importance of genetic variation for species, ecosystems, and
38 human society (Brown & Marshall 1995, Klein et al. 2009, though see Richards et al. 2007,
39 Schlottfeldt et al. 2015).

40 Genetic diversity is important because it allows species to adapt and survive under changing
41 conditions, and provides resilience to stresses such as drought, pests and diseases. Genetic
42 variation in foundation species, including trees, can stabilize ecosystems during environmental
43 fluctuations (Hughes et al. 2008). Where genetic diversity is lost, problems such as inbreeding
44 may occur. Furthermore, genetic diversity underlies traits which are valued by human societies
45 such as timber quality or disease resistance. Genetic diversity is therefore vital for artificial
46 selection or breeding of plants. The need to conserve genetic resources is well recognized in
47 the Convention on Biological Diversity (e.g. Aichi Target 13, <https://www.cbd.int/sp/targets/>). It is
48 particularly vital to understand, conserve, and make accessible forest genetic resources, as
49 recognized in international commitments like the Global Plan of Action on Forest Genetic
50 Resources (FAO 2014), the Montreal Process on sustainable forestry
51 (www.montrealprocess.org/), the Global Strategy for Plant Conservation
52 (<https://www.cbd.int/gspc/>), and a 2016 IUCN Resolution on autochthonous forest genetic
53 diversity (<https://portals.iucn.org/library/node/46521>).

54 Ex situ conservation (such as collecting seeds or plant parts for long-term storage and eventual
55 propagation) is an important component of strategies to conserve genetic diversity because in
56 situ populations experience a wide range of threats such as habitat change, over-exploitation,
57 climate change, invasive species and new pests (Cochrane et al. 2007, Li and Pritchard 2009,
58 Cavender et al. 2015). Well-planned seed bank collections play a vital complementary role to in
59 situ conservation by: preventing extinctions, providing seeds for species or habitat recovery, and
60 providing accessible, well-documented germplasm for research and experimentation. Seed
61 banks often prioritize the conservation of rare and threatened species, species required for
62 habitat restoration, or economically important crops and their relatives. It is particularly urgent
63 to bank seed of forestry species given their economic value, the ecological value of woodland
64 habitats, and the significant risk of catastrophic loss to disease or pests (FAO 2014).

65 The Millennium Seed Bank Partnership (MSBP), led by the Royal Botanic Gardens, Kew is the
66 largest ex situ conservation project in the world. The MSBP comprises a range of national and
67 international seed collecting projects that emphasize taxonomically verified, genetically

representative, geolocated seed of high quality and quantity, obtained with full consent and associated data (Liu et al. 2018). The MSBP UK National Tree Seed Project (NTSP), initiated in 2013, focuses on capturing infra-specific genetic diversity of target species within a politically-defined geographic area. UK trees and woodlands are threatened by an increasing number of exotic, problematic pests and pathogens, in addition to environmental uncertainties such as climate change. Most woodlands are small and isolated and have been modified or managed for centuries. In this context, the NTSP was developed to conserve the genetic diversity of UK tree species and facilitate their future survival. The project seeks to make genetically representative collections of UK native trees and shrubs, and to maintain these collections as a long term conservation strategy, but also to make them accessible for conservation and research activities. Collecting seed from a large number of mother trees, spread across a range of biogeographic zones, is expected to capture a high proportion of the genetic diversity present in the UK. A coordinated network (over 30 partner organisations, comprising over 600 professionals and citizen science volunteers) made collections from 2013- 2017 (Trivedi & Kallow, 2017; Kallow & Trivedi 2017, Trivedi 2016).

The NTSP provides a useful case study for common challenges to capture genetic diversity, whether for long-term conservation or more immediate use. Ex situ collections usually aim to best represent the genetic variation of the targeted population or species in order to be of the highest value to researchers or conservationists, though some collections target specific phenotypes or traits such as disease resistance rather than broad representation (such as whitebark pine in the USA, Mahalovich & Dickerson 2004). Genetically diverse collections will contribute to greater reintroduction success and provide a wide base for breeding or selection (Brown & Marshall 1995, Guerrant et al. 2004).

Genetically-based variation in form, phenology, pathogen resistance, environmental tolerance, and other traits are distributed heterogeneously among regions, populations and individuals. Thus a careful collection plan is needed that includes: where to collect, from how many individuals and/or populations to collect, and how much seed to collect. Existing guidelines (summarized in Hoban & Strand 2015) are based on simple mathematical equations (i.e. 50 individuals per population) first developed in the 1970s (Marshall & Brown 1975, Brown & Marshall 1995, Lawrence et al. 1995, Lockwood et al. 2007), or best guesses based on intuition (50 populations, Brown & Hardner 2000, or 5 populations, CPC 1991). A further recommendation is to select populations throughout the species' range, because range edge populations could have unique genetic variation (Lesica and Allendorf 1995). Some recommendations also include practical considerations, for example, to collect 10 times as much seed as the anticipated need, to allow for duplication of the collection for safeguarding, use of the collection for research, germination testing over time, sharing with other users, and other losses (Way 2003). The commonness of a species and the purpose of the ex situ collection are also used to determine the collection plan (Guerrant et al. 2013). For common species it is useful to have a larger collection (i.e. to use for restoration, research or distribution). Ideal recommendations must also be adjusted to practical limitations, such as when a population does not have 50 fruiting individuals at the time of collection, or when species are locally uncommon.

Ideally seed collections should be tailored to species' traits, their geographic distribution, and the planned end use of the seed, but quantitative guidance (e.g., numbers of populations, trees or seeds) is lacking (CPC 1991, Guerrant et al. 2004, 2013, Hoban et al. 2015). Evidence for the benefits of different sampling strategies is sparse and underdeveloped; as a consequence, most collecting protocols follow the "50 individuals per population minimum" but show wide variation in total number of seed, spatial sampling advice, and other considerations (Hoban and Strand 2015). Therefore, conservation collections may not be effectively or efficiently conserving genetic variation.

Recent progress has been made on developing two main quantitative approaches to designing seed collections for a target species (or evaluating the genetic diversity of seed collections which are already made): genetic surveys and simulation studies. (1) A genetic survey uses DNA markers to assess genetic diversity across a species' range. Computer algorithms are used to subsample this dataset to "try out" different sampling strategies and produce very specific advice for a target species (Griffith et al. 2014, 2017, McGlaughlin et al. 2015). This approach has the downside of requiring thorough collections of DNA from tissue samples across the species' range, in order to quantify existing in situ genetic diversity. This is not feasible for common species and is challenging and expensive even for rare species. (2) Mathematical or simulation models based on a species' traits such as dispersal characteristics or population structure can be used to produce broad guidelines for species having a given trait (Hoban and Strand 2015, Kashimshetty et al. 2017) and to identify general principles (Hoban and Schlarbaum 2014). The downside of this approach is that models used so far have been simple and the guidelines are not species-specific. Both approaches have been applied to relatively few cases.

We demonstrate here a new approach for designing seed collection strategies for target species, which combines these two approaches. It uses demographic-genetic models tailored to the target species: a model at fine-scale including a realistic simulation of the species' dispersal biology and a model at large scale including realistic spatial distribution of populations. The models are used to produce a synthetic estimate of a species' extant genetic diversity, and then to test different sampling strategies in order to identify effective strategies. The model can be checked against real genetic data from previous studies. We demonstrate that this approach can be used to produce sampling guidelines for a target species, without requiring an intensive genetic survey covering the whole range. This approach therefore solves the disadvantages of the two approaches mentioned above (expense and lack of specificity).

We focus our new approach on the case of European ash (*Fraxinus excelsior*), a common tree across Europe that contributes important ecosystem functions (Mitchell et al. 2014). The effective ex situ conservation of UK ash genetic diversity is a high priority for the NTSP because ash is the second most common native hardwood tree (126 million ash trees in the UK, making up 14% of forest standing volume) but is threatened by the devastating disease ash dieback (caused by a recently described ascomycete fungus, *Hymenoscyphus fraxineus*, previously *Chalara fraxinea*). This disease was first observed in Poland in 1992 and in the UK in 2012, and now exists in more than 60% of 10 km forest observation plots in England and Wales (<https://www.forestry.gov.uk/ashdieback>) and continues to rapidly spread (Pautasso et al. 2013, Lobo et al. 2014). While ash dieback does not kill all trees, mortality may reach 95%. It is also

recognized that in the future Emerald Ash Borer could reach the UK and provide an additional threat, making collections particularly urgent (Thomas 2016).

In this paper we explain our new method for designing national seed collections for a common species; use our new approach to evaluate the likely genetic capture of European ash collections made by NTSP thus far; and determine best practice guidelines (numbers of trees, number of populations, and spatial locations of populations) for future collections. We answer the following questions, while also analyzing results to draw broader conclusions for ex situ seed collections: Has the NTSP adequately captured the genetic diversity of ash in the UK by sampling 59 sites in 22 seed zones, 659 ash trees and 2.3 million seeds? How appropriate is current advice for collectors in the NTSP manual, including minimum sample sizes and recommendations for sampling across the range? What changes might be recommended for future sampling?

2 Methods

2.1 Species description

European ash (*Fraxinus excelsior* L.) occurs across Europe in mixed broadleaf forests or rarely in pure stands. It is especially common in the United Kingdom. Ash is valued economically for its flexible wood and ecologically as a foundation species. More than 1000 species are associated with ash and 44 are obligate dependents on ash (Thomas 2016). Ash seed and pollen are dispersed by wind, with pollen movement documented over 3 kilometers (Bacles et al. 2006, Bacles and Ennos 2008). Ash is a relatively early colonizing, large (typically 10-20 meters tall, up to 40 meters tall) tree, and lives to a maximum of approximately 250 years. Seeds are typically produced yearly but mast years occur approximately every 2-5 years (Thomas 2016). Ash seed collected by NTSP have a mean germination rate of 83%.

Genetic marker studies of ash reveal low population differentiation as expected for a high dispersal species (Heuertz et al. 2001, Ferrazzini et al. 2007, Sutherland et al. 2010). Common garden studies show differences among provenances planted in different locations, but differences in performance were not always clearly related to geography (Cundall et al. 2003). Local provenances often, but not always, outperformed non-local ones. Climatic gradients in the UK include an east-to-west gradient of dry-to-moist air, and north-to-south gradient of cool to warm. The south is low-lying, while the north has more hills and mountains (up to 1300 meters). Ash is common in the central and southern areas, and fragmented and rare in the north.

2.2 NTSP Sampling Strategy

The NTSP has developed sampling protocols at the national and local scale, with a goal of collecting a broad and representative base of genetic diversity for practical use and scientific study. Collectors receive a UKNTSP Collecting Manual with a collecting protocol, which is used across all target species (Kallow 2014).

National sampling strategy: At a national scale, the protocol states that a “species will be ‘fully represented’ when it has been collected from every seed zone in which it occurs or from all genetically-distinct populations where this is known.” The 24 seed zones are determined by the Forestry Commission based on biogeographic and human landscape features (Herbert et al. 1999). They are not species-specific (except for Scots Pine). Ash seed has been collected by the NTSP from all seed zones in which the species occurs. The NTSP has a secondary goal to collect (when possible) at least one high altitude (>300 meters) and one low altitude collection per seed zone in which the species occurs. It is worth noting that the word “population” is difficult to apply in common outcrossing trees with wind-dispersed pollen and long distance seed dispersal mechanisms, which often results in continuous genetic structure rather than distinct genetic populations (Craft & Ashley 2007, Hoban et al. 2010, Victory et al. 2006). The term Target Collection is therefore used to describe the collection of locally distinct collections.

Target Collection sampling strategy: the protocol details several components to ensure that each Target Collection is as genetically diverse and as representative of the autochthonous genes and local adaptations as possible. Site selection requires ancient, semi-natural, native woodlands to reduce the likelihood of sampling from planted, foreign trees. After site selection, the goal is to capture the maximum diversity from the trees in that woodland. Collectors are instructed to select trees randomly rather than choose trees with any particular traits. Selected trees should be at least 50 meters apart to avoid siblings. (This distance is based on previous studies. Semizer-Cumming (2017) fit dispersal kernels and estimated average distances for ash seed dispersal at 39 m, with 50% of seeds traveling less than 37 m. Heuertz (2003) suggest gene flow from seed dispersal at 14 m, and pollen at 70 m.) At least 15 trees should be sampled to make a Target Collection. This number is a compromise between widespread seed collecting guidance to collect from 50 trees and the time and human resources it takes to collect from trees (15 is the maximum number that can be sampled in one day by a team of 4). Notably, the UK native woodland is extremely fragmented, often into small woodlots; a Target Collection from 15 trees may require collections across several small woodlots or numerous lone trees.

Seed is to be taken across the tree canopy to ensure that the seed has been pollinated by paternal trees in all directions. The goal per Target Collection is 10,000 viable seed, a minimum that was chosen to allow sufficient seed for future use, duplication, and viability monitoring. To facilitate future research purposes (e.g., provenance trials, trait screening and heritability testing) collectors store seed from individual mother trees separately and mother trees are tagged and geo-referenced so that researchers can return to them.

Using this guide, the NTSP made 59 *Fraxinus excelsior* collections, from 54 ten km² areas, from 22 seed zones, for a total of 659 trees (mean of 11.2 trees per collection, and 29.3 trees per seed zone). (Note that there are 9 additional collections from 71 maternal trees in the Millennium Seed Bank from pre-NTSP collections, which are not considered here). The number of viable seed sampled per tree is approximately 3470.

2.3 Methods summary

Our new approach leverages the advantages of the two mentioned above: a detailed, species-specific genetic survey providing knowledge of the amount and distribution of a species' genetic diversity, and a trait-based model providing knowledge of how to sample due to a species' inherent characteristics. We use detailed models tailored to the ecology, genetics, and spatial distribution of ash (in contrast to general models used previously, Hoban & Schlarbaum 2014, Hoban & Strand 2015) which takes the place of and obviates the need for a full genetic survey across many thousands of trees. There are four steps (FIGURE 1). (1) Detailed genetic-demographic models are built using observed ash occurrence data from the Forestry Commission and information regarding ash ecology, genetics, and population biology. (2) The models are run, producing datasets of genetic diversity for the individuals and populations in the model. These genetic datasets are large (>one million individuals and >one thousand populations for the national scale model), far larger than any empirical genetic study, thus fully describing allele frequencies including rare alleles and considering all populations. The simulated data represent estimates of the genetic variation existing across GB and how genes are distributed among populations, trees and seeds. (3) The simulated genetic datasets are sub-sampled according to different sampling strategies (numbers of populations, individuals and seeds), including the NTSP sampled populations. (4) The genetic diversity captured in each proposed sampling strategy is compared to the genetic diversity (in terms of number of alleles) that exists in the full simulation; we calculate the proportion of all extant alleles that are actually sampled. This proportion is an estimate of how much existing genetic diversity (in the wild) is actually captured in the ex situ seed sample. Higher proportions mean more genetic variation conserved. We describe these steps below, with technical details in the Supplemental Materials. Advantages and disadvantages of this approach are provided in the Discussion.

Figure 1: Conceptual overview of new approach using simulations of ash in four steps: build species-specific models, run simulations to create genetic datasets representing ash populations, sample from these datasets, and quantify performance of the sampling

2.4 Steps 1 and 2: models and prior knowledge used to create genetic datasets

Our approach includes a large- and a small- spatial scale model (Figure 1). Simulations at two scales are necessary due to computational limits. It is impossible to model all seeds on all living ash trees (i.e., trillions of seeds). Modeling at two scales is also consistent with seed collecting guidelines, which have large- and small-scale protocols. Lastly, each scale addresses different questions. The large-scale model estimates genetic variation *among populations* and what could be captured in various strategies in terms of number of populations and trees, as well as geographic locations of chosen populations. The small-scale model estimates the genetic variation existing *within a population* and what could be captured with different strategies in terms of number of trees and seeds, and whether trees are sampled randomly or with spatial aggregation. Each model is described below. Our models assume that the main phenomena influencing distribution of genetic variation are migration and population size. More assumptions are included in the Supplemental Materials

Large-scale model: This model is used to determine how many and which populations to sample. This is a spatially realistic model covering the ash distribution using realistic estimates of population sizes and distribution. The large-scale model determines the spatial distribution of genetic diversity across GB and within each local cell or population. Genetic theory states that well-connected large populations should show higher genetic diversity, while small and disconnected ones will show lower, but perhaps unique, genetic diversity. Populations are arranged on a grid of cells, each 10 km on a side, and simulated with a coalescent algorithm using the software fastsimcoal and custom-written R scripts (see Supplemental Methods document). This scale roughly corresponds to maximum pollen movement (Bacles et al 2006) and thus genetic connectivity. The grid corresponds to the Ordnance Survey National Grid, which is also the scale of the National Forestry Inventory that we used for parameterizing our model, (see below) containing 9100 grid cells. A rectangle encompassing ash distribution is 4717 cells, with 1092 cells occupied (unoccupied cells are ocean or unsuitable habitat). Population sizes are parameterized using Forestry Inventory data (non-public dataset) that contains an estimate of ash proportion of the local canopy at 6918 plot locations across GB. Note that not every cell has an observation plot and some cells have multiple observations. Each population is considered panmictic in this model. We tested sensitivity of results to different population sizes and migration rates (see Supplemental Methods). Summary statistics (i.e., F_{ST} and heterozygosity) from simulation outputs were compared to previous genetic studies in *F. excelsior* to ensure realistic results (Heuertz et al 2001, Sutherland et al 2010; see Supplemental Methods).

Small scale model: This model is used to determine how a real sampling would perform locally. The small scale model occurs within each population and models each individual through its lifespan. It is informed by prior knowledge of plant reproduction and ecology (e.g. life stages, dispersal) from natural history studies. Each individual experiences a juvenile phase, matures into an adult via a stage transition matrix, receives pollen, and produces seed. All individuals are hermaphrodites, a simplification of true ash biology (Thomas 2016). We used a dispersal kernel consistent with seed and pollen observations from molecular genetic parentage studies (Bacles et al 2006, Bacles and Ennos 2008). Based on our initial simulations and work in other species (Gapare et al 2005, McGlaughlin et al 2015), we hypothesize that the same sampling effort performed in different local populations might capture different amounts of genetic variation, partly because different populations have different amounts of genetic variation to

start with (e.g., isolated vs. well-connected populations). We recognize that other forces influence genetic variation (i.e., climate, local inbreeding, rainfall, etc.), but here we simply assume that population size and connectivity are the main forces. We therefore simulated and sampled from 23 different local populations (chosen across the large simulation grid) that varied by population size, location in the range (south, north, edge, center) and proximity to other populations. This allows for comparing sampling strategies within edge vs. core populations, north vs. south, etc. (e.g., a smaller sampling effort may be sufficient within isolated, lowgenetic-variation populations). More details are in the Supplemental Methods document.

Additional simulation at European-scale: We primarily focus on sampling ash genetic diversity within GB. However, substantial genetic variation likely exists throughout Europe. Therefore we produce a simulation of ash at the European-scale, testing two possible bottleneck and two possible migration scenarios, to estimate how well the NTSP represents genetic diversity across Europe (see Supplemental Methods).

2.5 Step 3: Sampling of simulated datasets

Large-scale model: We tested sampling strategies that differed by number of trees per population, number of populations, and spatial arrangement of populations. We tested all combinations from 2 to 50 plants and 2 to 70 populations for each of the geographic sampling strategies in Figure 2, and the NTSP sampling described above. For NTSP sampling the number of plants was adjusted to account for multiple seeds taken (see Supplemental Methods and Supplemental Figure S2).

Figure 2: geographic sampling strategies tested, left to right, top to bottom: random draw from all populations; each population from a different latitudinal band (18 latitudinal bands, termed 'regions'); only populations from northern two bands; only populations from southern two bands; only populations from central two bands; only populations from core/center of the range (no edges sampled); only populations from margin/edge of the range (no central populations sampled); populations mostly (75%) from southern two bands with random sampling for the remaining 25% of populations; and populations mostly (75%) from northern two bands with random sampling for the remaining 25% of populations. These are termed: "random", "each from different region", "only northern 2 rows", "center 2 rows", "only southern 2 rows", "core", "edge", "focus south", "focus north." In addition, we sampled where the NTSP has already sampled (See Supplemental).

Small-scale model: We tested all combinations of 1 to 50 plants and 2 to 1500 seeds per plant, in each of the 23 populations simulated, for each of two spatial strategies: ‘random’ in which sampled trees are chosen randomly through the whole population and ‘clumped’ or aggregated, in which a sampler chooses the first n trees that s/he encounter near his/her starting point.

2.6 Step 4: Analysis

Usually conservation scientists do not know the functional or adaptive role of each allele (each gene can have several alleles, or variants), and often assume that every allele can have potential value. Thus we quantify how well the sampling captures allelic diversity. The simulation outputs a list of all alleles in GB, A_t , while the sampling step outputs a list of alleles occurring in the seed sample, A_s . Dividing A_s by A_t results in a “proportion of genetic variation captured” at least once in the sample, which is the main result we will consider. We calculate this proportion for all alleles and separately for categories of alleles based on commonness at a given geographic scale. For example an allele can be common (high frequency), intermediate, or rare (low frequency). Alleles that are common locally but rare elsewhere (i.e., at high frequencies in only one location), are particularly interesting as this is the pattern expected if the allele is locally adapted and has been preserved through natural selection. Capturing alleles with such a pattern may represent capturing locally adaptive alleles, though the simulation does not actually simulate the process of adaptation. Details including thresholds for allele categories are in the Supplemental Methods.

Additional interest in capturing multiple copies: The main results concern “the proportion of allelic variation captured.” Consistent with previous work on ex situ seed collections, if an allele is captured even once, that genetic variant is considered conserved. Success is defined as the presence of one copy in the collection. However, alleles present in only one seed will probably not survive through to an outplanted, restored population. Some seeds are inviable and some seedlings die, and alleles are lost during regeneration or breeding cycles (Basey et al. 2015). Data from Cochrane et al. (2007) analyzed a number of plant species and estimated that 7% to 76% of seeds (mean of 30%) will result in adult plants. Moreover, sub-samples of the collections will be made to many different researchers who will want to access these seeds. Therefore a better goal may be to capture many copies of each allele to allow for such losses and multiple uses. We therefore calculated “proportion of genetic variants captured fully” which requires collecting at least 50 copies of each allele.

3 RESULTS

3.1 NTSP performance

We estimate that NTSP collections have captured approximately 91% of all alleles, 100% of moderate to low frequency alleles, 77% of rare alleles and 96% of locally common alleles. Only 52% of locally rare alleles are captured.

3.2 Ranking of alternative sampling strategies- when to sample edge vs. center

Most of the following results will focus on “all alleles”, though some interesting results for other allele categories will be noted. On average, the best large-scale spatial strategies were random and stratified random, followed by edge, core, and south (Figure 3). Consistently the worst was to sample only in the north, followed by focus on the north, and only in the center. This result is consistent for other allele categories (i.e. rare, etc.)

Figure 3: Mean relative performance of each spatial sampling strategy for proportion of alleles captured, over a range of number of trees sampled

Interestingly, the relative ranking of sampling strategies changed depending on how many populations could be sampled (Supplemental Figure S3). If the sampler can only reach 2 or 5 populations, the best strategies are only south, focus on the south, and core. If the sampler can reach >5 populations, the best strategies are random, stratified random, and edge (followed by focus on the south and core). This result is consistent for other allele categories (i.e. rare, etc.).

3.4 Comparing NTSP actual sampling to other potential strategies

NTSP sampling performed nearly as well as random sampling. For “all alleles” (as well as low frequency, rare and locally rare alleles) the only spatial sampling strategies that could have performed better, with fewer total samples, was random with 55 populations and 10 trees

sampled per population (550 total trees vs. NTSP total of 659). Moreover, most other sampling strategies could never capture as much diversity as NTSP even if more trees and populations are visited than the NTSP strategy (Table 1 includes a partial list in order to illustrate that most strategies require many more samples than the NTSP in order to capture more genetic diversity than the NTSP sampling). Results are similar for “locally common” alleles, though somewhat greater sampling would be needed for south-focused sampling to exceed the NTSP performance for locally common alleles (Table 1).

All alleles (global)			Locally common alleles		
Sampling	Total plants sampled	Populations sampled from	Sampling	Total plants sampled	Populations sampled from
random	1125	25	random	1225	35
random	875	35	random	1125	45
random	550	55	random	825	55
each fr diff reg	1575	35	each fr diff reg	1800	45
each fr diff reg	1375	55	each fr diff reg	1650	55
each fr diff reg	975	65	each fr diff reg	1300	65
core	1225	35	core	NA	NA
edge	1575	35	edge	2000	50
edge	1125	45	edge	1650	55
edge	1100	55	edge	1400	70
focus S	1375	55	focus S	NA	NA
focus N	3250	65	focus N	NA	NA

Table 1: A partial list of sampling strategies that could achieve equivalent or higher genetic diversity capture than the actual NTSP sampling. Note that nearly all of these required more total sampled trees than employed with the NTSP (659 trees, 59 sites), emphasizing the good performance of the NTSP. For locally common alleles, core, focus north and focus south never exceeded the NTSP performance even with maximum number of samples tested (70 populations), thus NA is reported.

3.5 Choice of new population or new tree and points of diminishing returns

Examining genetic capture vs. number of trees and genetic capture vs. the number of populations (Figure 4), we observe that curves for number of populations are steeper and plateau later than for number of trees. In other words, sampling will capture more genetic

variation if it is spread among more sites (see examples below). The same conclusion is reached when examining graphs for the other allele types, except for locally rare, in which sampling more populations or more trees results in an equivalent outcome.

We then sought to determine when to stop sampling- usually a somewhat arbitrary task- when additional *gains* are below a threshold of *returns* (and thus, when to stop). Specifically, we calculated how much genetic variation (for all alleles) is gained at every sampling point on the curve when adding an additional unit (tree or population). We then identified the sampling point at which gain is less than 1% additional genetic diversity per added “unit of effort,” i.e. additional sampling is not gaining appreciable additional benefit. Using this threshold, the point of diminishing returns is 6.62 trees. The point of diminishing returns for a 0.5% threshold is 11.6 trees, and for a 0.1% threshold it is 28.9 trees. Results are similar for other allele types, though the point of diminishing returns (for a threshold of 0.5%) is higher for low frequency (e.g. 13.6 trees) and rare alleles (18.7 trees). The point of diminishing returns for populations ranges from 35 to 37 populations depending on the threshold, and is consistent for different allele types.

Figure 4: Accumulation curves showing genetic variation (y axis) captured for (at Left) adding more populations (x axis), where each successive line assumes a given number of trees sampled per population; and (at Right) adding more trees per population (x axis) where each successive line assumes a given number of populations sampled. Not all lines are labeled due to space

3.6 Capturing at least 50 copies of an allele

Fewer alleles are captured “fully” when using a threshold of 50 copies for alleles to be considered preserved. In this case NTSP has fully captured 72% of all alleles and 77% of locally common alleles. We then asked: “does the number of samples necessary scale with desired

number of copies e.g. does a collector need 50 times as many samples to capture 50 copies of each allele?" Remarkably, the answer is approximately "yes". Examining cases of equivalency, we observed that between 30 and 60 times as many samples (mean of 48) from 5 to 10 times as many populations (mean of 7.5) would be required to achieve a given proportion (e.g. preserving 70% of all alleles) if success required 50 copies per allele instead of one copy.

3.7 Sampling compared to the European range

Under all scenarios simulated, sampling within GB captures less genetic variation than sampling within the Balkans or Iberian Peninsulas, and much less than sampling across the whole European continent (Figure 5). In the low migration, and large or small bottleneck scenario (a likely historical event), GB sampling captures under 75% of alleles while sampling in a different European region captures over 90% of the variation and sampling throughout Europe captures nearly 100%. For high migration scenarios GB sampling improves to about 80% of alleles, but is still less than sampling any one other European region or sampling across Europe.

Figure 5: Relative performance of sampling within the UK vs. other regions of Europe vs. across Europe.

3.8 Sampling randomly vs. clustered within a population

For within population sampling, we observed that sampling randomly across the population always resulted in a higher proportion of genetic diversity (for all alleles) captured than clustered sampling- on average 0.05 greater (a relative increase of 8.7%). However, in some populations sampling randomly could be 17% better, though the difference depended on sampling effort.

The difference between random and clustered sampling was much higher when sampling from fewer trees (less than 20) and many seeds (more than 20) per tree (Figure 6). For such cases, random sampling could increase the proportion captured by 0.22 over clumped (a 47% increase). It was also observed that it is most important to sample randomly in the southern, well-connected populations (Supplemental Figure S5). Results differ somewhat among the allele types. For all allele types, random sampling is most important for small to moderate numbers of trees, while for low frequency alleles there is no effect of number of seeds.

Figure 6: The difference in genetic variation (in one example population) captured by sampling randomly and in clustered fashion, for different numbers of trees and seeds- the difference is greatest, and thus the importance of sampling randomly is highest, for sampling few trees and many seeds. Values are negative as they are calculated as “captured in clustered sampling” minus “captured in random sampling”.

3.9 Sampling differently within populations depending on location, size and connectivity.

We sampled 23 local populations differing in range location, population size, and connectivity to nearby populations. The genetic diversity in terms of all alleles (though results are qualitatively similar for low frequency alleles) captured from each population, on average and for particular sampling strategies, is shown in Figure 7. Isolation clearly influences proportion of genetic variation captured, with smaller influence of range location and population size. The mean genetic variation captured when sampling in the clumped fashion in a non-isolated population is 0.685 while sampling an isolated population is 0.845, a gain of 23.4% (0.750 and 0.869 when sampling randomly). This effect increases for smaller sampling efforts especially when sampling smaller numbers of trees (colored dots in Figure 7). Sampling from two trees with 200

seeds each (a realistic sampling) from a moderately connected, medium-to-large size population (population number 12) captured 0.373, while sampling in the same way in the isolated, northern, small population (population number 1) yielded 0.687- almost twice as much genetic diversity.

Figure 7: The genetic diversity (all alleles) captured in 23 local populations over all sampling efforts (grey boxplots) and for specific efforts (colored dots). Note that genetic capture is higher in isolated populations than in central populations for the same sampling effort. X axis abbreviation: N (north), MN (middle north), M (middle), C (core), S (south); min (minimally connected), isol (isolated), half (half populations connected), maj, (majority are connected); 1 to 7 are relative population size. Plot is shown for clumped sampling within populations but is similar for sampling randomly.

3.10 Choice of new tree or more seeds, including diminishing returns

We find that choosing another tree is much better than choosing another seed from already sampled trees (Figure 8), with rapidly diminishing returns when adding more seeds. For a threshold diminishing return (for all alleles) of 1% the stopping point is 19.1 trees and 8.09 seeds per tree, while for a threshold return of 0.5% it is 31.2 trees and 9.73 seeds and for a threshold return of 0.1% it is >50 trees and 30.9 seeds per tree (similar to the diminishing returns analysis at the range-wide scale).

Figure 8: Illustration of diminishing returns when collecting more seeds (brown) or collecting from more maternal trees (green). Note the strongly differing slopes. The left plot is the same data as the center plot but extended out to 1500 seeds. This plot shows results for all alleles for a central population. Patterns (i.e. rapid plateau when adding more seeds) are highly similar for other allele types and populations, though genetic diversity accumulates faster for low frequency than rare alleles and faster in northern populations.

4 DISCUSSION

We have presented a new approach to evaluating and choosing seed sampling strategies. This approach combines features of the general, trait-based simulation approach (e.g. Hoban and Schlarbaum 2014) and genetic survey approach (e.g. Griffith et al 2014). We developed a multi-scale, detailed simulation parametrized on European ash biology, existing detailed forest survey data, and a small existing genetic survey (Sutherland et al. 2010). We then calculated the genetic diversity likely captured by the actual NTSP sampling and various other possible sampling strategies. We show that for the common goal of capturing one copy of each allele, the NTSP has likely captured more than 90% of global and locally common alleles (see Supplemental for definitions of these), assuming that the focus is on neutral genetic diversity. We were able to test and identify optimal sampling solutions, both among and within populations, and to identify situations in which certain sampling strategies (e.g. random

sampling) are most important. We also showed that smaller sampling effort within populations is sufficient in northern isolated compared to southern core populations. Then, we quantified points of diminishing returns more precisely than in previous studies. Lastly, we evaluated capture for a conservation goal of 50 allele copies rather than one copy.

It is not surprising that random geographic choice of populations is best. The finding that south and core perform very well is likely because the species is densely populated in these areas. Good performance of edge sampling (i.e., sampling throughout the periphery) especially at higher population numbers is likely due to covering large geographic distances, and/or because edge populations have unique alleles. Theory predicts that alleles will drift to high frequencies in each edge population in a stochastic manner (Lesica and Allendorf 1995, Arenas et al. 2011), and thus different alleles will be at high frequencies in different edge populations while central populations mostly share similar common alleles. Random or widely distributed sampling seems to be especially important for capturing locally common alleles (Table 1). The NTSP strategy showed very good performance and is nearly random. Virtually no strategy could have done better with fewer samples. Interpreting these results, we make the following general advice: If sampling a moderate or large number of populations, cover the whole range, including range-edge populations; if sampling a small number of populations, focus primarily but not exclusively in the abundant core of the range. If future NTSP collections (from ash, and potentially other tree species) are seeking to add additional site locations to the existing seed collection, they would benefit from visiting additional edge populations.

It is clear from the European simulations that GB-based sampling can capture an appreciable but far from complete representation of the full species range; full capture will require sampling across the continent. These results are based on a very simple model as a first estimation, and a more sophisticated model is needed, but this general result should hold. Examination of several simulation files from our study revealed that GB does not have any private alleles, i.e. alleles only present in GB but not in Europe, while Europe typically has several alleles not present in GB. In other words the genetic diversity of GB is likely a subset of European diversity. Importantly though, we emphasize that this is neutral, not adaptive, genetic diversity.

At the local scale, we highlight the genetic inefficiency of sampling too many trees and too many seeds per tree. For the goal of one copy of each allele, diminishing returns are observed after about 30 populations, 10-30 trees and only about 10-30 seeds per tree, where less than 1% additional genetic variation is captured per added sampling unit. This means that additional sampling beyond this point captures extremely marginal gains in genetic diversity. Especially dramatic diminishing returns are seen for seeds per tree. For example, sampling 1000 seeds per tree barely captures more genetic diversity than sampling 100 seeds! The genetic capture from one tree with 1500 seeds is approximately equal to sampling between 5 and 30 trees (mean of 10) with one seed each, a total seedlot size at least 50 times smaller. The relevance of this finding, and whether to stop at the point of diminishing returns, will depend on the intended use of the seed collection. If the seed are for a genetic study or to establish a seed orchard or site planting then numbers of seed per tree should meet the requirement of those tasks. A general-purpose seed bank needs to serve multiple users over many years, which may warrant continuing to collect many seeds per tree. It remains to be seen whether these quantitative results will apply for all species, but they are consistent with previous general findings by simple

573 simulations (Hoban and Schlarbaum 2014) and the fact that typically few paternal parents (<30)
574 contribute to a maternal tree's seed set (Sork et al. 2002).

575 In some situations, the NTSP collections achieved fewer than 15 trees per site, as UK
576 woodlands are small and highly fragmented, and also due to the constraints of volunteer
577 collectors. Ideal NTSP efforts could improve sampling effectiveness by revisiting sites in which
578 fewer than 15 trees were sampled to ensure this minimum is achieved (after reviewing data
579 sheets to ensure there are additional local trees available), or visiting additional, new sites.

580 We also show that a given amount of genetic variation can be captured with a smaller sample
581 size in isolated northern populations than in the south, though the "correct" amount to sample
582 depends on the conservation target and nonlinear returns. The Figures should help collectors.
583 Our results are consistent with the work of Griffith et al. (2014) and McLaughlin et al. (2015) who
584 showed that different populations need to be sampled differently because some populations
585 have more genetic diversity, inbreeding, or connectivity than other populations.

586 Also at the local scale, we find that higher genetic capture is achieved when sampling randomly
587 (i.e. across a large landscape). However in contrast to previous work showing a large effect
588 (Hoban and Strand 2015), we find a relatively small difference on average (5 to 10%) between
589 random and clustered sampling for this species. Still, this difference is magnified in specific
590 cases: sampling few maternal trees (2 to 20) and/or many seeds per tree, and sampling in the
591 south, where the difference was 20 to 40%. In such cases it is particularly advised to choose
592 trees as far apart as possible, and thus the current NTSP protocol which advises random
593 sampling of trees within populations is sound. A smaller effect than in previous studies (Hoban
594 and Strand 2015) is likely due to the relatively long pollen and seed dispersal of ash.

595 Sampling strategies are rarely, if ever, evaluated for a goal of capturing multiple copies of each
596 allele. This is important because not all seeds will survive to an adult plant and as such some of
597 the individuals and their alleles will be lost. Additionally, for some collections, seeds may be
598 distributed to many users and each user will want copies of every allele. We showed that if the
599 goal is to capture 50 copies of each allele, NTSP has likely only captured approximately 70% of
600 alleles being "fully preserved," which is a less-optimistic appraisal of conservation success. We
601 further showed that a collection whose goal is 50 copies of each allele should sample
602 approximately 50 times as much seed (from approximately 7 times as many populations) as a
603 collection aiming to capture one copy of each allele. This finding supports large seed banks for
604 common species. It remains to be seen if a simple multiplication could be a general rule for
605 collectors wanting multiple allele copies (i.e. would sampling 5 times as much result in 5 copies
606 of each allele), but we predict it will. In any case, we emphasize that a deep level of duplication
607 is desirable for ecologically important collections like ash, and should be planned for, especially
608 in light of known threats from pests and diseases. If NTSP aims for 50 copies of each allele,
609 about 150 to 250 "sites" (individual populations) need to be sampled, each with 15 to 25 trees,
610 and 1000 seeds per tree (which is 5 times as many sites and 25 to 30 times as many total trees
611 sampled as needed if only one copy was required, which is consistent with the multiplication
612 rule presented above; total of 2.3 million to 6.3 million seeds). Interestingly, while these
613 numbers seem very large, the number collected by NTSP so far (2.3 million seeds) has reached
614 the lower end of this range, but it would be better if the seeds were distributed among many

more trees and sites. Such large collections may or may not be feasible, but this is what would be needed for 50 copies of the alleles.

It is worth noting that the same proportion of genetic capture can be reached with different sampling strategies, i.e. different *means* can be used to reach the same *ends*. We call these “genetically equivalent collections”. For instance in Table 1, we observe that a collector could focus on the south and achieve the same genetic capture as sampling randomly, but they would have to sample more than 4 times as many trees per population for locally common alleles or 2 times as many trees per population for all alleles. There are similar genetic equivalences for the choice of number of trees vs. populations. For example, a sampler can capture 71% of all alleles by sampling 5 populations and 50 trees per population, or by sampling 45 populations with 2 trees per population. The latter strategy has more logistical effort (having to reach many more populations) but an overall sample size less than half as large as the former (90 samples instead of 250), and the collections have similar genetic variation. Another example: 85% can be captured by sampling 45 populations with 15 trees each (675 samples), or 20 populations with 50 trees each (1000 samples). Sometimes the optimal solution will not be possible for a collector, but this does not mean the collection should not go forward! For example, it may only be possible to sample in some regions or populations due to political or logistical constraints, or it may be impossible to sample randomly at the local scale due to time constraints or property boundaries. A good genetic capture can still be made, but the sampler should adjust their protocol using the genetic equivalence concept (for example to sample more trees and/or seeds per tree than if they could follow an ideal protocol). The phenomenon of genetically equivalent collections has not been explicitly discussed in the literature but has great practical use. We emphasize that the previous standard “50 samples” is not best for everyone; 15 plants sampled, with many more seeds per plant, can be good (though sampling <5 plants is never good). As we did not test every possible sampling strategy, or the interaction of every sampling choice, further investigation is advised.

A critique of this approach might be that the results we find are dependent on the assumptions and information behind models we’ve used. For example, if our migration model is incorrect, our estimates of allele capture (e.g. overall 90 to 95% of alleles captured with one copy) are over-optimistic and real levels may be lower, because lower levels of migration would result in more genetic variation distributed among populations and thus a poorer performance of sampling. However, lower migration rates would result in genetic differentiation much higher than are observed in ash populations based on genetic studies in this and similar species (Heuertz et al. 2001, Sutherland et al. 2010, Flower et al. 2018), and we have likely chosen the most realistic parameters. More importantly, the general patterns we observe do not change, i.e. the relative rankings of each strategy are the same, even if the exact percentage of capture differs under different migration assumptions (Supplemental Figure S1). Our broad conclusions are likely robust to possible limitations in our model. We note that our model does not consider possible large-scale human influences such as plantings. Ash is a commercial species with some introductions of non-local material in recent decades. However, the NTSP project has only collected from undisturbed, non-planted, ancient woodlands (Rackam 2014), using data sources such as the Ancient Woodland Inventory (<http://lle.gov.wales/catalogue/item/AncientWoodlandInventory2011/?lang=en>) or the Native Woodland Survey of Scotland (<https://scotland.forestry.gov.uk/supporting/strategy-policy->

guidance/native-woodland-survey-of-scotland-nwss). Such locations should have preserved the native genetic constitution of the species. We note also that a parentage analysis of ash in a fragmented landscape in Germany (Simizer-Cumming et al. 2017) found the majority of seeds and seedlings to have local parents. Sutherland et al (2010) found genetic evidence of isolation by distance which is a phenomenon expected under natural gene flow rather than human movement of genes. In any case, for future application of this approach in other species, it is possible to incorporate knowledge of planting history into the migration model.

There is also a philosophical assumption to note, which must be kept in mind for interpreting all results: that genetic variation from neutral mutations is the desired outcome of a seed collection. There are relationships between neutral genetic variation, fitness, variation at adaptive markers, and species survival (Reed & Frankham 2003, Leimu et al. 2006), but the genetic basis of adaptation is complex and difficult to identify with current methods (Hoban et al. 2016, Flanagan et al. 2018). For example, species with low genetic variation can sometimes show high phenotypic variation and an ability to respond to selection (Roman & Darling 2007). However, scientists currently have limited knowledge on the traits and genes that will be important for species in the future, so we treat this subject consistently with the traditional philosophy of ex situ collections: higher genetic diversity will provide option value even if we do not currently understand the function or use of that diversity (Sarr et al. 2008). We do not know how much genetic variation a species needs, and so we seek a maximum amount until a reasonable point of diminishing returns. Neutral genetic markers also do not represent traits or co-adapted gene complexes. We note that provenance trials can reveal important geographic variation in genetically based traits, but these studies are available for few species and are not generally used to design conservation seed collections. Future research is needed to determine whether and how provenance results could help design seed collections.

While we have used a common (though sometimes locally uncommon, depending on local forest composition) species of conservation concern for our case study, our approach should be feasible for any species for which sufficient knowledge of biology and distribution is available; minimum data would include range maps, occurrence points, and pollination and seed dispersal mechanisms. However, as noted in the Introduction, the abundance of the focal species and purpose of the collection are important to consider. For instance, the approach that we use here is not necessary for very rare species if every flowering individual can be sampled, or if only one population exists. The approach does not require a genetic survey, although genetic data in similar species will help to set realistic model parameters. While we use this approach for ex situ conservation, it could be used for in situ conservation, such as in situ conservation areas designated to conserve genetic diversity, where decisions must be made on the number of reproductive individuals to include in each unit, total number of units to be designated per area, and location of the units in the species' range.

5 SUMMARY

In summary our new, detailed simulation approach to designing seed collections, using a mix of data and models, is advantageous because it provides full estimation of a species' genetic variation and allows testing of strategies at multiple scales without the cost and challenges of a detailed genetic survey. This approach has allowed a detailed assessment of the conservation

success of NTSP ex situ ash collections, and has provided a better understanding of limitations and possible changes to the NTSP sampling strategy. The ability to quantify the genetic outcomes of different sampling strategies can provide scientific justification to seed collecting programmes, reassurance to their funders, and much needed practical guidance for field collectors. Some of our results (i.e. diminishing returns after 30 populations) are specific to the ash system, but other results (i.e. sampling in edge vs. core populations) should be universal. Furthermore, the approach might also provide a robust means to measure and monitor the success of nations or organizations in achieving stated ex situ conservation objectives, i.e. to measure progress against Convention on Biological Diversity Aichi Target 13 (Hollingsworth, pers comm). Also this study illustrates the importance of considering the number of trees sampled as well as the number and geographic arrangement of populations sampled when developing indicators of success for ex situ targets. Overall, we predict that our approach will enable more efficient and effective conservation of species' genetic resources, ultimately providing ecological and societal benefits.

COMPETING INTEREST STATEMENT: The authors declare no competing interests or conflicts of interest.

FUNDING: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors (though see Acknowledgements).

REFERENCES

- Arenas, M., Ray, N., Currat, M., Excoffier, L. 2011. Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution*. 29(1):207-218.
- Bacles, C.F., Lowe, A.J., Ennos, R.A. 2006. Effective seed dispersal across a fragmented landscape. *Science*. 311(5761):628-628.
- Bacles, C.F.E., Ennos, R.A. 2008. Paternity analysis of pollen-mediated gene flow for *Fraxinus excelsior* L. in a chronically fragmented landscape. *Heredity*. 101(4):368.
- Basey, A.C., Fant, J.B., Kramer, A.T. 2015. Producing native plant materials for restoration: 10 rules to collect and maintain genetic diversity. *Native Plants Journal*. 16(1):37-53.
- Brown, A.H.D., Hardner, C.M. 2000. Sampling the gene pools of forest trees for ex situ conservation. *Forest Conservation Genetics: Principles and Practice*. Melbourne (Australia): CSIRO. pp.185-196.
- Brown, A.H.D., Marshall, D.R. 1995. A basic sampling strategy: theory and practice. *Collecting Plant Genetic Diversity: Technical Guidelines*. Wallingford: CAB International. p.75-91.

735 Boyce, M.S., 1992. Population viability analysis. *Annual Review of Ecology and Systematics*.
736 23(1):481-497.

737 Carwardine, J., Klein, C.J., Wilson, K.A., Pressey, R.L., Possingham, H.P. 2009. Hitting the
738 target and missing the point: target-based conservation planning in context. *Conservation*
739 *Letters*. 2(1):4-11.

740 Cavender, N., Westwood, M., Bechtoldt, C., Donnelly, G., Oldfield, S., Gardner, M., Rae,
741 D., McNamara, W. 2015. Strengthening the conservation value of ex situ tree collections. *Oryx*.
742 49(03):416-424.

743 Cochrane, J.A., Crawford, A.D., Monks, L.T. 2007. The significance of ex situ seed conservation
744 to reintroduction of threatened plants. *Australian Journal of Botany*. 55(3):356-361.

745 Craft, K.J., Ashley, M.V. 2007. Landscape genetic structure of bur oak (*Quercus macrocarpa*)
746 savannas in Illinois. *Forest Ecology and Management*. 239(1-3):13-20.

747 Cundall, E.P., Cahalan, C.M., Connolly, T. 2003. Early results of ash (*Fraxinus excelsior* L.)
748 provenance trials at sites in England and Wales. *Forestry*. 76(4): 385-400.

749 Food and Agriculture Organization (FAO). 2014. Global Plan of Action for the conservation,
750 sustainable use, and development of forest genetic resources. Rome: Commission on Genetic
751 Resources for Food and Agriculture

752 Ferrazzini, D., Monteleone, I., Belletti, P. 2007. Genetic variability and divergence among Italian
753 populations of common ash (*Fraxinus excelsior* L.). *Annals of Forest Science*. 64(2):159-168.

754 Flanagan, S.P., Forester, B.R., Latch, E.K., Aitken, S.N. and Hoban, S., 2017. Guidelines for
755 planning genomic assessment and monitoring of locally adaptive variation to inform species
756 conservation. *Evolutionary Applications*.

757 Flower, C.E., Fant, J.B., Hoban, S.M., Knight, K.S., Steger, L., Aubihl, E., Gonzalez-Meler, M.A.,
758 Forry, S., Hille, A., Royo, A.A. 2018 Apr 13. Optimizing conservation strategies for a threatened
759 tree species: in situ conservation of white ash (*Fraxinus americana* L.) genetic diversity through
760 insecticide treatment. *Forests*. 9(4):202.

761 Forestry Commission, 1998. The UK Forestry Standard: the Government's approach to
762 sustainable forestry.

763 Gapare, W.J., Aitken, S.N., Ritland, C.E. 2005. Genetic diversity of core and peripheral Sitka
764 spruce (*Picea sitchensis* (Bong.) Carr) populations: implications for conservation of widespread
765 species. *Biological Conservation*. 123(1):113-123.

766 Griffith, M.P., Calonje, M., Meerow, A.W., Tut, F., Kramer, A.T., Hird, A., Magellan, T.M., Husby,
 767 C.E. 2014. Can a botanic garden cycad collection capture the genetic diversity in a wild
 768 population? *International Journal of Plant Sciences*. 176(1):1-0.

769 Griffith, M.P., Calonje, M., Meerow, A.W., Francisco-Ortega, J., Knowles, L., Aguilar, R., Tut, F.,
 770 Sánchez, V., Meyer, A., Noblick, L.R., Magellan, T.M. 2017. Will the same ex situ protocols give
 771 similar results for closely related species? *Biodiversity and Conservation*. 26(12):2951-66.

772 Guerrant, E.O., Jr., Havens, K., Vitt, P. 2013. Sampling for effective ex situ plant conservation.
 773 *International Journal of Plant Sciences*. 175(1):11-20.

774 Guerrant, E.O., Jr., Havens, K., Maunder, M. eds. 2004. *ExSitu Plant Conservation: Supporting*
 775 *Species Survival in the Wild (Vol. 3)*. Island Press.

776 Herbert, R., Samuel, S., Patterson, G. 1999 Aug. *Using Local Stock for Planting Native Trees*
 777 *and Shrubs*. Edinburgh: UK Forestry Commission.

778 Heuertz, M., Hausman, J.F., Hardy, O.J., Vendramin, G.G., Frascaria-Lacoste, N., Vekemans, X.
 779 2004. Nuclear microsatellites reveal contrasting patterns of genetic structure between western
 780 and southeastern European populations of the common ash (*Fraxinus excelsior* L). *Evolution*.
 781 58(5):976-988.

782 Heuertz, M., Vekemans, X., Hausman, J.F., Palada, M. and Hardy, O.J., 2003. Estimating seed
 783 vs. pollen dispersal from spatial genetic structure in the common ash. *Molecular Ecology*, 12(9),
 784 pp.2483-2495.

785 Heuertz, M., Hausman, J.F., Tsvetkov, I., Frascaria-Lacoste, N., Vekemans, X., 2001.
 786 Assessment of genetic structure within and among Bulgarian populations of the common ash
 787 (*Fraxinus excelsior* L.). *Molecular Ecology*. 10(7):1615-1623.

788 Hoban, S.M., 2014. An overview of the utility of population simulation software in molecular
 789 ecology. *Molecular Ecology*. 23(10):2383-2401.

790 Hoban, S., Kelley, J.L., Lotterhos, K.E., Antolin, M.F., Bradburd, G., Lowry, D.B., Poss, M.L.,
 791 Reed, L.K., Storfer, A. and Whitlock, M.C., 2016. Finding the genomic basis of local adaptation:
 792 pitfalls, practical solutions, and future directions. *The American Naturalist*, 188(4), pp.379-397.

793 Hoban, S.M., Strand, A. 2015. Ex situ seed collections will benefit from considering spatial
 794 sampling design and species' reproductive biology. *Biological Conservation*. 187:182-191.

795 Hoban, S.M., Borkowski, D.S., Brosi, S.L., McCleary, T.I.M., Thompson, L.M., McLachlan, J.S.,
 796 Pereira, M.A., Schlarbaum, S.E., Romero-Severson, J. 2010. Range-wide distribution of genetic
 797 diversity in the North American tree *Juglans cinerea*: a product of range shifts, not ecological
 798 marginality or recent population decline. *Molecular Ecology*. 19(22):4876-4891.

799 Hoban, S.M., Strand, A., Fraga, N., Richards, C.,Schlarbaum, S. 2015. Developing quantitative
800 seed sampling protocols using simulations: A reply to comments from Guja et al. and Guerrant
801 et al. *Biological Conservation* (184):469-470.

802 Hoban, S.M., Schlarbaum, S. 2014. Optimal sampling of seeds from plant populations for ex-
803 situ conservation of genetic biodiversity, considering realistic population structure. *Biological*
804 *Conservation*. 177:90-99.

805 Hughes, A.R., Inouye, B.D., Johnson, M.T., Underwood, N.,Vellend, M. 2008. Ecological
806 consequences of genetic diversity. *Ecology Letters*. 11(6):609-623.

807 Kallow, S., Trivedi, C. 2017. Collecting genetic variation on a small island. In *Proceedings of*
808 *Workshop on Gene Conservation of Tree Species*, May 16-19, Eds. RA Snieszko, R.A., Man, G.,
809 Hipkins, V., Woeste, K., Gwaze, D., Kliejunas, J.T., McTeague, B.A. Gen. Tech. Rep. PNW-
810 GTR-963. U.S. Department of Agriculture, Forest Service.

811 Kallow, S. 2014. *UK National Tree Seed Project Seed Collecting Manual*. Royal Botanic
812 Gardens, Kew. Available online:
813 <https://www.kew.org/sites/default/files/UK%20national%20tree%20seed%20project.pdf>

814 Kashimshetty, Y., Pelikan, S.,Rogstad, S.H. 2017. Effective seed harvesting strategies for the
815 ex situ genetic diversity conservation of rare tropical tree populations. *Biodiversity and*
816 *Conservation*. 26(6):1311-1331.

817 Klein, C., Wilson, K., Watts, M., Stein, J., Berry, S., Carwardine, J., Smith, M.S., Mackey, B.,
818 Possingham, H. 2009. Incorporating ecological and evolutionary processes into
819 continental-scale conservation planning. *Ecological Applications*. 19(1):206-217.

820 Lawrence, M.J., Marshall, D.F., Davies, P. 1995. Genetics of genetic conservation. I. Sample
821 size when collecting germplasm. *Euphytica*. 84:89–99

822 Leimu, R., Mutikainen, P.I.A., Koricheva, J., Fischer, M. 2006. How general are positive
823 relationships between plant population size, fitness and genetic variation? *Journal of Ecology*.
824 94(5):942-952.

825 Lesica, P., Allendorf, F.W. 1995. When are peripheral populations valuable for conservation?
826 *Conservation Biology*. 9(4):753-760.

827 Li, D.Z., Pritchard, H.W. 2009. The science and economics of ex situ plant conservation. *Trends*
828 *in Plant Science*. 14(11):614-621.

829 Liu, U., Breman, E., Cossu, T.A. and Kenney, S., 2018. The conservation value of germplasm
830 stored at the Millennium Seed Bank, Royal Botanic Gardens, Kew, UK. *Biodiversity and*
831 *Conservation*, pp.1-40.

832 Lobo, A., Hansen, J.K., McKinney, L.V., Nielsen, L.R., Kjær, E.D. 2014. Genetic variation in
833 dieback resistance: growth and survival of *Fraxinus excelsior* under the influence of
834 *Hymenoscyphus pseudoalbidus*. *Scandinavian Journal of Forest Research*. 29(6):519-526.

835 Lockwood, D.R., Richards, C.M., Volk, G.M. 2007. Probabilistic models for collecting genetic
836 diversity: comparisons, caveats, and limitations. *Crop Science*. 47:861–868

837 Marshall, D.R. and Brown, A.H.D. 1975. Optimum sampling strategies in genetic conservation.
838 In '*Crop genetic resources for today and tomorrow*'. (Eds OH Frankel, JG Hawkes) pp. 53–70.

839 Mahalovich MF, Dickerson GA. 2004. Whitebark pine genetic restoration program for the
840 Intermountain West (United States). *Breeding and genetic resources of five-needle pines:
841 growth, adaptability and pest resistance*. Proc RMRS-P-32, USDA For Serv Rocky Mountain
842 Res Sta, Ft. Collins, CO. Jul:181-7.

843 McGlaughlin, M.E., Riley, L., Brandsrud, M., Arcibal, E., Helenurm, M.K., Helenurm, K. 2015.
844 How much is enough? Minimum sampling intensity required to capture extant genetic diversity
845 in ex situ seed collections: examples from the endangered plant *Sibara filifolia* (Brassicaceae).
846 *Conservation Genetics*. 16(2):253-66.

847 Mitchell, R.J., Beaton, J.K., Bellamy, P.E., Broome, A., Chetcuti, J., Eaton, S., Ellis, C.J.,
848 Gimona, A., Harmer, R., Hester, A.J., Hewison, R.L. 2014. Ash dieback in the UK: a review of
849 the ecological and conservation implications and potential management options. *Biological
850 Conservation*. 175:95-109.

851 Pautasso, M., Aas, G., Queloz, V., Holdenrieder, O. 2013. European ash (*Fraxinus excelsior*)
852 dieback—a conservation biology challenge. *Biological Conservation*. 158:37-49.

853 Reed, D.H., Frankham, R. 2003. Correlation between fitness and genetic diversity.
854 *Conservation Biology*. 17(1):230-237.

855 Richards, C.M., Antolin, M.F., Reilley, A., Poole, J., Walters, C. 2007. Capturing genetic
856 diversity of wild populations for ex situ conservation: Texas wild rice (*Zizania texana*) as a
857 model. *Genetic Resources and Crop Evolution*. 54(4):837-48.

858 Roman, J., Darling, J.A. 2007. Paradox lost: genetic diversity and the success of aquatic
859 invasions. *Trends in Ecology & Evolution*. 22(9):454-464.

860 Sarr, M., Goeschl, T., Swanson, T. 2008. The value of conserving genetic resources for R&D: A
861 survey. *Ecological Economics*. 67(2):184-193.

862 Schlottfeldt, S., Walter, M.E.M., de Carvalho, A.C.P., Soares, T.N., Telles, M.P., Loyola, R.D.,
863 Diniz-Filho, J.A.F. 2015. Multi-objective optimization for plant germplasm collection conservation
864 of genetic resources based on molecular variability. *Tree Genetics & Genomes*. 11(2):16.

865 Semizer-Cuming, D., Kjær, E.D. and Finkeldey, R., 2017. Gene flow of common ash (*Fraxinus*
866 *excelsior* L.) in a fragmented landscape. *PloS one*, 12(10), p.e0186757.

867 Sork, V.L., Davis, F.W., Smouse, P.E., Apsit, V.J., Dyer, R.J., Fernandez-M, J.F., Kuhn, B.
868 2002. Pollen movement in declining populations of California Valley oak, *Quercus lobata*: where
869 have all the fathers gone? *Molecular Ecology*. 11(9):1657-1668.

870 Sutherland, B.G., Belaj, A., Nier, S., Cottrell, J.E., P Vaughan, S., Hubert, J., Russell, K. 2010.
871 Molecular biodiversity and population structure in common ash (*Fraxinus excelsior* L.) in Britain:
872 implications for conservation. *Molecular Ecology*. 19(11):2196-211.

873 Thomas, P.A. 2016. Biological flora of the british isles: *fraxinus excelsior*. *Journal of Ecology*.
874 104(4):1158-1209.

875 Trivedi, C. 2016. Banking on the future. Woodwise: Seeds of Hope. *Woodland Conservation*
876 *News*. Autumn:10-12.

877 Trivedi, C., Kallow, S. 2017. Benefits and challenges for gene conservation: a view from The UK
878 National Tree Seed Project. In *Proceedings of Workshop on Gene Conservation of Tree*
879 *Species*, May 16-19, Eds. Snieszko, R.A., Man, G., Hipkins, V., Woeste, K., Gwaze, D.,
880 Kliejunas, J.T., McTeague. B.A. Gen. Tech. Rep. PNW-GTR-963. U.S. Department of
881 Agriculture, Forest Service.

882 Victory, E.R., Glaubitz, J.C., Rhodes, O.E., Woeste, K.E. 2006. Genetic homogeneity in *Juglans*
883 *nigra* (Juglandaceae) at nuclear microsatellites. *American Journal of Botany*. 93(1):118-126.

884 Way, M.J. 2003. 'Chapter 9: Collecting seed from non-domesticated plants for long-term
885 conservation' in Smith, Roger D. *Seed Conservation: Turning Science into Practice*. London:
886 Royal Botanic Gardens, Kew. Pp. 165-201.

887

888

889