The Association between Mating System and Range and Genome Size within the UK Flora

Samuel Rawlinson

The University of Edinburgh

Evolutionary Biology Honours

26/04/2021

**CONTENTS**

[1 ABSTRACT 2](#_Toc70336443)

[2 INTRODUCTION 3](#_Toc70336444)

[3 MATERIALS AND METHODS 8](#_Toc70336445)

[3.1 The UK Flora Database 8](#_Toc70336446)

[3.2 Cleaning and Adding Data to the UK Flora Database 8](#_Toc70336447)

[3.3 Genome Size Analysis 9](#_Toc70336448)

[3.4 Range Size Analysis 11](#_Toc70336449)

[3.5 Phylogenetically Independent Contrasts 13](#_Toc70336450)

[3.6 Continent Presence 14](#_Toc70336451)

[3.7 Graphs 15](#_Toc70336452)

[4 RESULTS 16](#_Toc70336453)

[4.1 Genome Size 16](#_Toc70336454)

[4.2 Range Size 17](#_Toc70336455)

[5 DISCUSSION 25](#_Toc70336456)

[5.1 Genome Size 25](#_Toc70336457)

[5.2 Phylogenetic Generalized Linear Models 27](#_Toc70336458)

[5.3 Range Size 28](#_Toc70336459)

[5.4 Range Size and Life History 29](#_Toc70336460)

[5.5 Potential Confounding Factors in Range Size 30](#_Toc70336461)

[5.6 Theoretical Explanations for a Smaller Range Size in Selfing Species 31](#_Toc70336462)

[5.7 Relationships with Range versus Genome Size 32](#_Toc70336463)

[5.8 Implications for Theories Involving Selfing 33](#_Toc70336464)

[6 ACKNOWLEDGEMENTS 35](#_Toc70336465)

[7 REFERENCES 36](#_Toc70336466)

# ABSTRACT

Species which generally reproduce via self-fertilization, a mating system known as “selfing” in plants, are predicted to suffer from strong inbreeding depression. Understanding the interaction between selfing and other traits, such as genome and range size, may aide in understanding why selfing evolves despite these costs. Selection against selfish genetic elements, such as transposable elements which are known to be linked to genome expansion, is predicted to be stronger in selfing species and therefore may cause genome shrinkage. Meanwhile, selfing species are predicted to colonise new areas more easily and experience less gene flow of locally maladaptive alleles between populations so may be more efficient at expanding their ranges and have larger range sizes. These predictions were tested within the UK flora, using mating system (split into “selfing”, “mixed” and “outcrossing”) to describe the degree to which a species self-fertilizes. Differences in range and genome size between mating systems were tested and phylogenetic generalized linear models were fitted when these were significant to ensure differences were not due to shared phylogenetic history between species. No significant difference in genome size was found between mating systems. The effect of mating system on range size was found to interact with life history within the UK flora and outcrossing species were found to have significantly larger range sizes than selfing species on average within perennials, but no difference was found in annuals. The effect of mating system within perennials was the opposite to that predicted, suggesting other metapopulation processes or interactions between mating system and other traits may be more dominant determinants of range size within the UK than those predicted to result in selfing species having a larger range size.

# INTRODUCTION

Self-fertilization, or “selfing”, is when a sexually reproducing organism contributes both gametes to the production of offspring. It is known to occur across sexually reproducing taxa, including in both animals (Jarne and Auld, 2006) and plants (Goodwillie *et al.*, 2005). In plants, the degree to which a species self-fertilizes is used to define its mating system, with plants generally being classified into “selfing”, “mixed” and “outcrossing”. Selfing plants sexually reproduce predominantly or exclusively via self-fertilization. Outcrossing plants sexually reproduce predominantly or exclusively via “outcrossing”, which is when the gametes contributing to the production of offspring are contributed by two different individuals. Plants with a mixed mating system sexually reproduce via both self-fertilization and outcrossing.

However, the prevalence of selfing presents a puzzle. Self-fertilization is an extreme form of inbreeding, which many organisms have mechanisms to avoid due to the fitness costs of breeding between related individuals known as “inbreeding depression”. Inbreeding depression is thought to arise from breeding between individuals which share alleles due to shared heritage causing increased homozygosity in inbred offspring. This results in the loss of the fitness benefits of heterozygosity at loci where the heterozygotes have the fittest genotype (heterozygous advantage) and an unmasking of deleterious recessive alleles (reviewed in Charlesworth and Willis, 2009). Therefore, either the fitness benefits of selfing must be greater than the costs of inbreeding depression or a species must evolve mechanisms to avoid said depression for selfing to be an evolutionarily viable strategy. For some organisms, such as sessile organisms with a low population density, the benefit of selfing may be reproductive assurance when finding a mate can be challenging (Morris *et al.*, 2016). Alternatively, the increased homozygosity resulting from selfing may be advantageous in stable environments (reviewed in Otto, 2009). In such a scenario, the breaking up of adaptive genetic backgrounds due to recombination during sexual reproduction may be detrimental. The increased homozygosity caused by inbreeding may therefore be beneficial as it will mean more recombination occurs between the same genetic background and recombination will be less likely to break up beneficial genetic backgrounds. Also, although selfing organisms would likely initially suffer badly from inbreeding depression, the exposure of deleterious recessive alleles within a population will expose them to selection and allow them to be more efficiently purged from a population (Wright *et al.*, 2008). This would result in long-term selfing populations having a smaller load of deleterious alleles than a comparable outcrossing population.

Multiple traits are known to be associated with selfing. For example, so-called “selfing syndrome” is a set of traits including changes to the relative investment of hermaphroditic organisms into male and female function or the loss of traits associated with mate or pollinator attraction such as large colourful flowers in flowering plants (Sincard and Lenhard, 2011; see Fig. 1). There are three main reasons a characteristic may be associated with selfing. Firstly, the trait may drive the evolution of selfing, such as organisms with a low population density taking part in self-fertilization to assure reproduction when finding a mate can be difficult. Secondly, a trait may facilitate or be required for the transition to selfing. This would include the loss of mechanisms to prevent selfing such as self-incompatibility loci or the spatial separation of stigma and anthers in plants. Finally, the evolution of a trait may be driven by selfing, such as a lessened need for mate finding due to selfing resulting in the loss of traits associated with mate finding. The traits associated with selfing are not limited to those associated with sexual reproduction (such as investment in male and female function or mate attraction) and include other extrinsic and intrinsic factors such as range size and genome size, respectively. Understanding the association between selfing and extrinsic factors is important as it may allow better understanding of what environmental or ecological conditions may drive selfing, whilst intrinsic factors, especially genomic factors, may provide an insight into how selfing organisms avoid the negative effects of inbreeding.

|  |
| --- |
| A picture containing plant, flower, indoor, orchid  Description automatically generated |
| Fig. 1- Selfing syndrome is exemplified in the *Capsella* genus, where the generally selfing species *C. rubella* (right) has flowers with smaller petals and stigma and anthers which are closer together than those of the generally outcrossing species *C. grandiflora* (left) (Eckardt, 2011). From Eckardt, 2011. |

One trait which has a reasonably well-studied relationship with selfing is range size. For example, a recent meta-analysis study, including samples of plants native to every continent except Antarctica, found that selfing plants tend to have larger range sizes (Grossenbacher *et al.*, 2015). The capacity of a species to self-fertilize has also been observed to be associated with larger range size in *Collinsia* species, where obligate outcrossing *Collinsia* species (those with self-incompatibility mechanisms) have smaller range sizes than self-compatible species (Randle *et al.*, 2009). This may be due to links between mating system and colonisation. Plants with the capacity to self-fertilize are thought to be better able to colonise new areas as they are not limited by the need to find a mate or by the presence of pollinators, so two individuals are not required to form a founding population. Self-fertilization could also allow rapid population growth despite initially small populations or low population densities, leaving selfing species at less risk of extinction due to random stochastic events. Therefore, founding populations of selfing species during colonisation may be more likely to persist. This is exemplified by Baker’s Law, that selfing species are overrepresented in islands compared to continents (Stebbins, 1957), which is supported by self-incompatibility mechanisms which prevent self-fertilization being more common in continental than island plants (Grossenbacher *et al.*, 2017). This colonisation hypothesis comes under the umbrella of the metapopulation hypothesis of range size, that mechanisms which determine population dynamics also determine range size in plants (Hanski, 1998). Another theoretical reason for selfing species having a larger range under the metapopulation hypothesis relates to range expansion. During range expansion, peripheral populations of a sexually reproducing species will experience gene flow from populations at the centre of the species’ range. Alleles which are adaptive at the centre of a species’ range may be maladaptive to the conditions experienced by peripheral populations, therefore the constant re-introduction of them to peripheral populations may limit these populations’ ability to expand into areas with different ecological conditions (Kirkpatrick and Barton, 1997). However, unlike outcrossing species, selfing species will experience little or no gene flow between central and peripheral populations, allowing peripheral populations to adapt to novel conditions and expand their range without the constant re-introduction of maladaptive alleles from the central populations.

However, mating system is also known to interact with other characteristics which may be associated with range size. For example, selfing is known to be associated with reduced genetic diversity (Hamrick and Godt, 1996), which may be associated with reduced niche breadth and less capacity to adapt to novel environments during colonisation of new areas. According to the niche breadth hypothesis (Brown, 1984), plant range sizes are determined by their ecological niches, with plants with broader ecological niches and the ability to cope with a broader range of environmental and ecological conditions being predicted to have larger range sizes. Therefore, selfing species would be predicted to have smaller range sizes under this hypothesis. The relative importance of the metapopulation and niche breadth hypotheses may therefore modulate the direction and strength of the effect of mating system on range size.

Genome size may also be associated with mating system. Some studies, such as one into the Veronicaceae family (Albach and Grielhuber, 2004), have found that selfing taxa have smaller genomes than their outcrossing relatives. Roessler *et al.*’s (2019)experimental study into the effect of selfing on genome size, in which 11 outbred maize lines were self-fertilized to produce homozygous lines which were then self-fertilized for 6 generations, found that 3 inbred lines showed a decrease in genome size. This decrease was associated with a reduction in the transposable element (TE) content of their genome. This finding is supported by much theoretical work which suggests that there should be stronger selection against selfish genetic elements, such as TEs (Charlesworth and Langley, 1986; Morgan, 2001) and B chromosomes (Burt and Trivers, 1998) in selfing species. As such elements are known to drive genome expansion (Lee and Kim, 2014), selection against them would be expected to result in a reduction in genome size. However, the predicted smaller genome size of selfing species has not been upheld by all studies. For example, the self-compatible species *Capsella rubella*, which has the capacity to self-fertilize, has a larger genome size than its self-incompatible relative *C. grandiflora* (Ågren *et al.*, 2014). However, the reverse is true for *C. grandiflora* and another self-compatible species *C. ortientalis* (Ågren *et al.*, 2014) and self-compatibility is not informative about the degree to which a species self-fertilizes, only that it has the capacity to do so, so the degree of self-fertilization may be relevant to its effect on genome size.

Understanding the association between selfing and other traits may be informative as to how selfing evolves and is maintained despite its apparent costs. Investigation of extrinsic factors such as range size or ecological factors may be informative about the conditions under which selfing is more prevalent and allow the formation of hypotheses of the conditions which drive selfing, whilst intrinsic factors such as genome size or genomic TE content may provide an insight into the potential benefits of selfing and how inbreeding depression is avoided in selfing species. Exploring how the relative strength of the association of extrinsic and intrinsic factors may also provide some idea about the relative importance of ecological advantages of selfing versus genomic mechanisms to avoid the costs of inbreeding in the evolution of selfing.

This project explored whether range size and genome size show associations with self-fertilization in the UK flora. The UK flora provides a good dataset for such a study as it is well-studied and the species that make it up are known. Furthermore, these species tend to be well-researched, with their range, ecological niches and history (e.g., whether they are native, archeophytes introduced thousands of years ago or neophytes introduced in recent history) being well understood, allowing other potentially confounding factors to be accounted for when investigating said associations. The UK flora is also relatively young, with much of it being assembled since the end of the last glacial maximum, and is known to be species-poor when compared to mainland Europe (Turrill, 1948). As the British Isles have been an island since the submerging of Doggerland, Baker’s Law may be expected to result in an over-representation of selfing species within the UK flora, which may allow easier identification of any associations between selfing and other traits of interest.

This study aimed to determine whether there are any associations between range size and genome size, respectively, and selfing in the UK flora. I hoped to answer the questions:

* Is genome size smaller in selfing taxa within the UK Flora?
* Is there a relationship between range size and mating system within the UK Flora? If so, do selfing taxa tend to have a larger or a smaller range size than other mating systems?
* Is the association between mating system and genome size or mating system and range size stronger within the UK Flora?

# MATERIALS AND METHODS

## The UK Flora Database

This project is based on a database, hereafter referred to as the UK Flora Database, collated and partially cleaned by Sean O’Brien. The UK Flora Database is available at: <https://github.com/mrseanman/NERCflora>. This database is based on data collected from the Ecological Flora Database (Fitter and Peat, 1994) using web page scraping code. Additional entries were added from the PLANTATT database (Hill *et al.*, 2004) using the same technique.

The mating system entries for species within the UK Flora Database were cleaned by inferring mating system from other traits. For example, species which were dioecious or self-incompatible have no capacity to self-fertilize so can be inferred to be strictly outcrossing. Information from the Genetic Flora Database (Ruhsam *et al.*, 2020) and PLANTATT database was also used to add mating system entries to species which lacked them.

The UK Flora Database contains entries on multiple plant characteristics including mating system, genome size and range size. Mating system information, within the “myFertGen” column, is derived from the Ecological Flora Database with entries cleaned by O’Brien. Genome size data, within the “X2C.DNA.content..pg.” column, is also derived from scraping of the Ecological Flora Database but was not cleaned by O’Brien. Finally, the range size information, which was sourced from the PLANTATT database and is contained in the “myPlantAtRange” column, refers to the number of 10km2 squares of the British Isles a plant is present in.

All datasets produced and analyses carried out during this project are available at: <https://github.com/Rawls29/Honours-Project>.

## Cleaning and Adding Data to the UK Flora Database

The 2C DNA content entries of the UK Flora Database required pre-cleaning before they could be used in data analysis. As part of this, three new columns were added to the database: “my2CDNA”, “my2CDNA\_VAR” and “myPloidy”. Firstly, for all entries which contained values which were repeats of the same number within the “X2C.DNA.content..pg.” column, this number was recorded without repetition in the “my2CDNA” column. Entries with multiple values which were not repeats of the same number were cleaned with reference to the C-Value DNA database(Leitch *et al*, 2019). Searches were carried out for each species with multiple entries using the default search settings except for C-value being set to “2C (pg)”. If the C-value database only returned one entry and this matched a value for the species in the UK Flora Database, this value was used for the 2C DNA content entry in the “my2CDNA” column. If the C-value database search only returned one entry but it did not match a value for the species in the UK Flora Database, the source of the entry in the C-value database was consulted to verify the value displayed in the C-value database was correct and applied to the 2C value, and then this value was used for the 2C DNA content of a species over any prior values in the UK Flora Database. If the C-value database search returned multiple entries, the sources for these entries were consulted to confirm which ploidy level they applied to. If each entry applied to a different ploidy level, separate entries were created for each ploidy level of the species, with the ploidy an entry was referring to being included in the “myPloidy” column. To prevent duplication of species in later investigations into factors other than 2C DNA content, new rows for different ploidy levels only contained the species name and relevant 2C DNA entries in the three new columns. In some cases, the sources did not provide ploidy information for all the entries and established trends, such as tetraploids having roughly double the 2C DNA content of diploids, were used to make assumptions about the ploidy level a value referred to. If different entries were not referring to different ploidy levels and were confirmed to be accurate following consultation of the sources, the mean of these values was used for the entry of the “my2CDNA” column, and the variance of the values was recorded in the “my2CDNA\_VAR” column. These new columns were used over the old “X2C.DNA.content..pg.” column for all analyses involving 2C DNA content.

To allow for the investigation of trends within individual families, an entry on which family each species belonged to was added based on their assigned genus for the 76% of entries that lacked a family entry within the UK Flora Database. The families each genus belonged to were identified by using the genus name as the search term in the Plants of the World Online search tool (POWO, 2019) and filtering for genus entries only. The family a genus belonged to was then sourced from the entry on said genus. If the genus name had been historically assigned to multiple genera, the correct family was confirmed by searching for the first species belonging to the genus in the UK Flora Database when alphabetically ordered and filtering for species entries only and the family which that representative species was assigned to was used for the complete genus.

## Genome Size Analysis

All analyses for both genome size and range size were conducted in R version4.0 (R Core Team, 2020) using RStudio Version 1.4.1103 (RStudio Team, 2020).

The UK Flora Dataset contained 515 entries with both 2C DNA content and mating system information. These were all for species of flowering plants, 513 angiosperms representing 82 families and 2 gymnosperms from 2 families, *Juniperus communis* from the Cupressaceae family and *Taxus baccata* from the Taxaceae family. This set of entries will be referred to as the genome dataset.

The genome dataset was first subjected to a Shapiro-Wilks test to determine whether it was normally distributed and a Levene test to determine whether the variances of the mating system groups were homogenous using the shapiro.test function and the leveneTest function within the car library (Fox and Weisberg, 2019), respectively. The results were used to determine whether a parametric analysis would be appropriate or whether non-parametric analyses should be carried out instead. When appropriate, parametric analyses were favoured due to their higher statistical power. The data was initially non-normally distributed, but the variances did not differ significantly between the groups. As the data was non-normally distributed, logarithmic, square root and box-cox (using the MASS library in R; Venables and Ripley, 2002)transformations were attempted and Shapiro-Wilks tests, Levene tests and diagnostic plots were used to determine whether the assumptions of normality and homogenous variances between groups were met after transformation. As the data remained non-normally distributed following all transformations attempted, the data was subjected to a Kruskal-Wallis test to determine whether the median 2C DNA content differed between mating systems using the kruskal.test function in R.

The relationship between 2C DNA content and mating system was also tested in 7 families which had more than 20 species in the genome dataset. The normality and homogeneity of variances was tested in each family when non-transformed and logarithmically transformed using a Shapiro-Wilks test and Levene test, respectively. 5 families met these assumptions when logarithmically transformed so were subjected to an analysis of variance (ANOVA) test using the aov function in R. The remaining 2 were non-normally distributed regardless of transformation but met the assumption of homogenous variances when non-transformed so the non-transformed data was subjected to a Kruskal-Wallis test.

To investigate the effect of a transition in mating system on genome size, pairs of closely related selfing and outcrossing species were identified as each pair would likely represent a single transition in mating system. 15 genera within the genome dataset were identified which contained both selfing and outcrossing taxa. Three of these genera, *Briza*, *Callitriche* and *Orobanche*, only contained 2 species, one of which was a selfing species and the other an outcrossing one. A literature search was carried out into the remaining 12 genera to identify the relationships between the selfing and outcrossing species and to determine whether the genus contained a pair of closely related species with opposite mating systems. Such pairs were found in 7 genera, *Cephalanthera* (Moscone *et al.*, 2007), *Juncus* (Drábková *et al.*, 2006), *Luzula* (Drábková *et al.*, 2006), *Papaver* (Carolan *et al.*, 2006), *Senecio* (Pelser *et al.*, 2002), *Trifolium* (Ellison *et al.*, 2006) and *Veronica* (Taskova *et al.*, 2004), with *Trifolium* containing 3 pairs. *Cardamine* (Carlsen *et al.*, 2009) and *Lathyrus* (Asmussen and Liston, 1998) were found to not contain an appropriate pair. The relationships between the species of interest were not resolved for three genera, *Poa*, *Ranunculus* and *Vicia*. The difference between the genome size of the selfing and outcrossing species in each pair was then calculated and a two-tailed one sample t-test with µ=0 was run to test whether the mean difference within pairs differed significantly from 0. Ratios of genome size between the species within each pair were calculated by dividing the larger genome size by the smaller genome size to allow inferences about whether the species within a pair were likely to have different ploidy levels.

## Range Size Analysis

The UK Flora Database contained 986 entries with both range size and mating system information. All entries were for flowering plants, with 983 entries on angiosperms from 93 different families and 3 entries on gymnosperms from 3 families (*Juniperus communis*, *Picea sitchensis* and *Taxus baccata* from Cupressaceae, Pinaceae and Taxaceae, respectively). This set of entries will be referred to as the range dataset.

Shapiro-Wilks tests and Levene tests were used to assess whether the range dataset and any subsequently tested subsets were normally distributed and had homogenous variances between groups when grouped by mating system, respectively. Where appropriate, transformations were attempted to achieve these assumptions. As no transformation attempted was found to achieve a normal distribution of data in any subset, any transformation necessary to achieve homogeneity of variances between groups was carried out. A Kruskal-Wallis test was carried out to determine whether the median range size differed between mating system groups and a post hoc Pairwise Wilcoxon Rank Sum test was carried out to assess which groups differed when the Kruskal-Wallis test returned a significant result.

Following the first analysis of the range size dataset, outliers whose range was likely to be completely or mainly derived from introductions were excluded as well as one endemic apomitic hybrid (*Sorbus arranensis*) upon the advice of Dr Kidner. The analysis was repeated.

The analysis was then repeated in all families with over 30 entries in the range dataset and at least 8 outcrossing and 8 selfing species. These cut-offs were chosen to ensure that the independently analysed family would have a reasonably good representation of species from each mating system and therefore would be less likely to return a significant result due to one extreme outlier. Entries excluded from the second range dataset analysis were also excluded from this analysis.

Analysis of the range dataset outliers revealed an over-representation of alpine species. Due to the geography of the UK, alpine species would be expected to have a lower maximum range size than most groups in the UK flora as they are largely limited to mountainous regions of Scotland and Wales and the North Pennines in England. Coastal and aquatic taxa would also be expected to have different maximum range sizes due to it being defined by either the area of coastline or water for these taxa, both of which differ from the area of land in the UK. Therefore, these three groups were segregated out into subsets for independent analysis. To achieve this, the Online Atlas of the British and Irish Flora (BRC, 2020) entry for each species in the range dataset was consulted. Aquatic species were defined as those which grew exclusively in water according to the “Ecology” section of their Online Plant Atlas of the British and Irish Flora entry. Coastal species were defined as those whose range was predominantly or exclusively confined to the coast and whose “Ecology” section referred to coastal features (e.g., littoral, sand dunes or sea cliffs). Alpine species were defined as those whose “World Distribution” was defined as “Montane” or, in the case of endemics, whose “Ecology” section referenced “montane” or “alpine”. Analysis was carried out on each of these subsets as well as the remaining main range dataset.

The main range dataset was then subcategorised into native and non-native species using the “Status” entry for species within the UK Flora Database and both subcategories were analysed in the same way as before. This was done as non-native species may be expected to show unusual patterns in range size unrelated to mating system (e.g., a species having an unusually large range size due to introduction as an ornamental to parkland across the British Isles or the range size of a species not being affected by its mating system due to it being non-reproductive or producing solely vegetatively within the British Isles). No distinction was made between neophytes and archeophytes as both included species which occurred in the UK purely through causal occurrences and were not naturalised according to the Online Plant Atlas of the British and Irish Flora.

The native subset was then further subcategorised based on life history as an exemplar trait of mature plants. This was done as Thompson and Ceriani (2003)suggested that the traits of mature plants were important in determining the range size of plants within the UK. The native subset was subdivided into an “annual” subset, which was all species which were defined as annual by their “myPlantAtLife1” entry in the UK Flora Database, and a “perennial” subset, which was all species not defined as annual by their “myPlantAtLife1” entry. The perennial subset therefore contained both biennial and perennial species. The annual and perennial subsets were then subjected to analysis as before.

The perennial subset was then re-analysed following the removal of three endemic *Sorbus* species, *S. arranensis*, *S. porrigentiformis* and *S. devoniensis*, which were removed as they are all endemic hybrids (Chester *et al.*, 2007; Robertson *et al.*, 2003) with unusually small range sizes (e.g., *S. arranensis* is confined to the isle of Arran whilst *S. porrigentiformis* is predominantly found in the Great Orme area of North Wales; BRC, 2020). The perennial subset was re-analysed again following the removal of phanerophytes due to the post-hoc hypothesis that outcrossing range size may have been artificially inflated by trees (which may have larger range sizes due to planting for forestry, for example) which are predominantly outcrossing (only 5 species of native tree are generally selfing, and of these only 2, *Sorbus aria* and *S. torminalis*, had not previously been excluded as hybrid endemics).

## Phylogenetically Independent Contrasts

To correct for phylogenetic relatedness between species and ensure phylogenetic non-independence between species was not responsible for significant differences between groups, phylogenetic generalized linear models (pgls models) were fitted for the groups which had shown significant differences in prior analyses: aquatic plants, native plants and perennial natives. Pgls models were also fitted for the complete genome dataset.

Phylogenies were produced for the genome dataset and aquatic and native range size subsets (see Appendix). Phylogenies were produced to the familial level with reference to a poster summarising the findings of the Angiosperm Phylogeny Group IV (Byng *et al.*, 2018). Polytomies were used at the base of a family to link all genera belonging to that family. Polytomies were also used at the base of a genus to link all species belonging to that genus. This was done as the relationship between genera and species (e.g. between and within the *Juncus* and *Luzula* genera; Drábková *et al.*, 2006) is not yet definitive for all taxa in the UK flora. These phylogenies were written in the Newick format and imported into R using the ape library (Paradis and Schliep, 2019).

Pgls models were fitted for all groups using the pgls function in the caper library (Orme *et al.*, 2018) within R with kappa, lambda and delta set to “ML”. The equation “(transformed) range/genome size ~ mating system” was used for these models. Diagnostic plots were consulted to confirm the assumptions of normality and homogenous variances were met. Logarithmic transformations of data were attempted to see if they improved agreement with assumptions. Logarithmic transformation improved agreement within aquatic plants, but not natives or perennials or within the genome dataset.

A pgls model was fitted for the native dataset which included mating system, life form and life history as additive factors and an interaction term between mating system and life history using the equation “range size ~ mating system + life form + life history + mating system\*life history”. This was done to confirm whether mating system has a significant effect on range size when variation due to life form and life history was controlled for. Kappa, lambda and delta were set as “ML”. A pgls model was also fitted for the perennial dataset which included mating system and life form as additive factors using the equation “range size ~ mating system + life form” and with kappa, lambda and delta set as “ML”. Interactions involving life form were not fitted as error messages occurred when this was attempted. Each model was run multiple times, with different factors fitted last, to obtain statistical values from type 3 sums of squares. This ensured there was no undue effect of the order of fitting of values. If this were not done, the variables fitted first would be more likely to be returned as statistically significant as any variation shared between the variables would be attributed to that which was fitted first.

## Continent Presence

To confirm whether range size within the UK is reflective of global range size, the number of continents species within the native range size subset were present on was determined with reference to the Global Biodiversity Information Facility(GBIF; GBIF.org, 2021). Occurrences were searched for by species name and the number of continents a species is present on was scored based on the map of georeferenced records. When a continent had few occurrences, the occurrences for the area were checked to confirm the validity of the species presence. If all the entries for a continent were from old observations (before 1990), referred only to planted or cultivated specimens, were occurrences where presence was inferred from one specimen, were all from publicly created datasets such as iNaturalist which may therefore be more prone to misidentification or may refer to non-wild specimens, or had ambiguity in the species classification, then the continent was not included in the count. For the purposes of distinguishing between species endemic only to Europe and those spread across the whole continental land mass of Eurasia, Europe and Asia were defined as separate continents with the classical borders of the Ural and Caucasus mountain ranges used.

Box-cox transformed range data was fitted against continental presence counts in a linear model using the lm function in R, with range size as the response variable and continental presence counts as the explanatory variable. Generalized linear models were also attempted using the glm function in R, but they did not improve meeting of the assumptions of the model. Linear models were fitted for range versus continent presence data for both the native and the perennial subset, with diagnostic plots being consulted to assess whether the assumptions of normality and homogenous variances had been met.

## Graphs

All graphs were produced using the ggplot2 (Wickham, 2016) and beeswarm (Eklund, 2016) libraries within R.

# RESULTS

## Genome Size

Median 2C DNA content did not vary significantly between mating systems within the UK flora (Kruskal-Wallis test, Χ2= 2.82, p>0.05; see Fig. 2). Average 2C DNA content was also not significantly different between mating systems within the Poaceae (ANOVA, F2/53= 0.046, p>0.05), Asteraceae (Kruskal-Wallis test, Χ2= 0.137, p>0.05), Fabaceae (Kruskal-Wallis test, Χ2= 0.896, p>0.05), Brassicaceae (ANOVA, F2/28= 2.09, p>0.05), Ranunculaceae (ANOVA, F2/20= 1.59, p>0.05), Plantaginaceae (ANOVA, F2/19= 3.45, p>0.05) or Rosaceae (ANOVA, F2/19= 0.667, p>0.05) families.

|  |
| --- |
| Chart, box and whisker chart  Description automatically generated |
| Fig. 2- 2C DNA content (pg) of species within the UK flora when grouped by mating system (generally cross n= 244, mixed n= 174, generally self n= 97). |

Furthermore, mating system did not significantly predict variation in 2C DNA content following correction for phylogenetic relatedness (pgls, F2/472= 1.60, p>0.05).

Selfing species had a smaller 2C DNA content than outcrossing species in 8 out of 12 species pairs (see Fig. 3). However, the difference in 2C DNA content did not differ significantly from 0 (one sample t-test, t=-1.11, p=0.29). The ratio of smaller to larger genome size within these pairs was roughly equal to or greater than 2 for 8 out of 12 pairs.

|  |
| --- |
| \*  \*  \*  \*  \*  \*  \*  \* |
| Fig. 3- Difference in 2C DNA content (pg) between pairs of related selfing and outcrossing species. The selfing species is labelled first and the outcrossing species second on the x axis. Positive values represent the selfing species having a larger 2C DNA content, and negative values represent the outcrossing species having a larger 2C DNA content. Asterisks (\*) denote pairs where the ratio between the smaller and larger 2C DNA content within a pair is roughly equal to or greater than 2. |

## Range Size

Median range size differed significantly between mating systems within the UK Flora (Kruskal-Wallis test, Χ2= 14.8, p= 0.00060), with outcrossing species tending to have a larger range size than species with a mixed mating system (Pairwise Wilcoxon Rank Sum test, cross\*mix, p=0.00046) and selfing species (Pairwise Wilcoxon Rank Sum test, cross\*self, p=0.044).

Outcrossing taxa also tended to have a larger range size at the familial level. Outcrossing species had a larger average range size than selfing species in 19 out of 27 families which contained both selfing and outcrossing taxa (see Fig. 4). 3 out of 7 families tested independently showed significant differences in median range size: Poaceae (Kruskal-Wallis test, Χ2= 8.29, p= 0.016), Plantaginaceae (Kruskal-Wallis test, Χ2= 8.06, p= 0.018) and Rosaceae (Kruskal-Wallis test, Χ2= 10.8, p= 0.0043). Outcrossing species tended to have a larger range size than selfing species in Poaceae (Pairwise Wilcoxon Rank Sum test, cross \*self, p= 0.015) and Plantaginaceae (Pairwise Wilcoxon Rank Sum test, cross\*self, p= 0.00026). However, within Rosaceae selfing species tended to have a smaller range size than species with a mixed mating system (Pairwise Wilcoxon Rank Sum test, cross\*mixed, p= 0.0041) but there was not a significant difference between selfing and outcrossing species (Pairwise Wilcoxon Rank Sum test, cross\*self, p>0.05).

|  |
| --- |
| Chart  Description automatically generated |
| Fig. 4- Difference between the mean range size of selfing and outcrossing species within 27 families which contained at least one selfing and outcrossing species, respectively, within the range dataset from the UK Flora Database. Positive values represent the selfing species having a larger range size, and negative values represent the outcrossing species having a larger range size. |

Following the removal of taxa which were expected to have low maximum range sizes, there was still a significant difference in median range size between mating systems (Kruskal-Wallis test, Χ2= 9.42, p=0.0090) with outcrossing species tending to have a larger range size than selfing species (Pairwise Wilcoxon Rank Sum test, cross\*self, p= 0.0056).

Non-native species were not responsible for this trend as there was no significant difference in range size between mating systems in non-native taxa (Kruskal-Wallis test, Χ2= 1.39, p=0.50; see Fig. 5a). The trend observed in the overall range dataset was also seen when only native species were analysed, with significant differences in median range size between

|  |  |
| --- | --- |
| a) | Chart  Description automatically generated |
| b) |  |
| Fig. 5- Range size of the a) native and b) non-native species within the UK flora when grouped by mating system (native generally cross n= 279, mixed n= 288, generally self n= 157; non-native generally cross n= 37, mixed n= 31, generally self n= 13). | |

mating systems (Kruskal-Wallis test, Χ2= 9.98, p=0.0068) and native outcrossing taxa tending to have larger range sizes than selfing taxa (Pairwise Wilcoxon Rank Sum test, cross\*self, p=0.0056) (see Fig. 5b). Mating system explained a significant amount of variation in range size in native taxa when fitted as the sole explanatory variable (pgls, F2/661= 5.51, p= 0.0042) but ceased to be significant when life history and life form were also included as explanatory variables (pgls, F2/647= 2.47, p>0.05). However, there was a significant interaction between mating system and life history (pgls, F4/647=2.94, p=0.020), as well as a significant effect of life form (pgls, F8/647= 4.20, p= 0.044).

Independent analysis of native annuals and native perennials suggests mating system has a larger effect in perennials. This is because native annual taxa showed no significant difference in range size between mating systems (Kruskal-Wallis test, Χ2= 4.54, p>0.05) but range size differed significantly between mating systems in native perennial taxa (Kruskal-Wallis test, Χ2= 9.98, p=0.0068) (see Fig. 6). Outcrossing perennials tended to have a larger range size than selfing species (Pairwise Wilcoxon Rank Sum test, cross\*self, p=0.044) and species with a mixed mating system (Pairwise Wilcoxon Rank Sum test, self\*mixed, p=0.044). This was maintained following the removal of 3 endemic species: *Sorbus arranensis*, *S. porrigentiformis* and *S. devoniensis* (Kruskal-Wallis test, Χ2= 8.26, p=0.016; Pairwise Wilcoxon Rank Sum test, cross\*self, p=0.024, cross\*mix, p=0.024). No significant difference in range size between mating systems was found when phanerophytes were removed from the perennial dataset, but this result was borderline (Kruskal-Wallis test, Χ2= 5.95, p=0.051). Both mating system and life form were found to have significant effects on range size within perennials following correction for phylogeny (pgls, mating system F2/457= 3.56, p= 0.029; life form F6/457= 4.75, p= 0.00010).

Within aquatic plants, median range size differed significantly between mating systems (Kruskal-Wallis test, Χ2= 15.2, p=0.00042; see Fig. 7a). Outcrossing species tended to have larger range sizes than both species with a mixed mating system (Pairwise Wilcoxon Rank Sum test, cross\*mix, p=0.018) and selfing (Pairwise Wilcoxon Rank Sum test, cross\*self, p=0.00093) species. However, mating system did not explain significant variation in range size (pgls, F2/53= 2.11, p>0.05).

Median range size did not differ significantly between mating system groups in either alpine (Kruskal-Wallis test, Χ2= 2.80, p>0.05; see Fig. 7b) or coastal (Kruskal-Wallis test, Χ2= 5.57, p>0.05; see Fig. 7c) taxa.

|  |  |
| --- | --- |
| a) |  |
| b) |  |
| Fig. 6- Range size of the a) native annual and ) native perennial species within the UK flora when grouped by mating system (annual generally cross n= 39, mixed n= 59, generally self n= 97; perennial generally cross n= 240, mixed n= 169, generally self n= 60). | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| a) | Chart, box and whisker chart  Description automatically generated | b) | Chart  Description automatically generated | c) | Chart, box and whisker chart  Description automatically generated |
| Fig. 7- Range size of the a) aquatic, b) alpine and c) coastal species within the UK flora when grouped by mating system (aquatic generally cross n= 28, mixed n= 10, generally self n= 20; alpine generally cross n= 43, mixed n= 41, generally self n= 23; coastal generally cross n= 99, mixed n= 74, generally self n= 69). | | | | | | |

**Global Distribution**

Range size within the UK was positively correlated with the number of continents a species is present on in both native taxa (linear model, F1/662= 185, p< 2.2e-16; see Fig. 8) and native perennials (linear model, F1/467= 176, p< 2.2e-16).

|  |
| --- |
| Diagram  Description automatically generated |
| Fig. 8- Range size of the native species within the UK flora grouped by the number of continents they are present on (species present on 1 continent= 42, species present on 2 continents= 71, species present on 3 continents= 113, species present on 4 continents= 108, species present on 5 continents= 92, species present on 6 continents= 238). The fitted line represents a linear model of box-cox transformed range size fitted against the number of continents a species was present on. |

# DISCUSSION

## Genome Size

No relationship was found between genome size and mating system in the complete flora dataset before or after phylogenetic correction or within any family tested. There was also no significant difference between the genome size of closely related selfing and outcrossing pairs.

The lack of a significant relationship within any analysis of genome size shows that mating system does not have a detectable relationship with genome size within the UK Flora. Therefore, selfing species were not found to have a smaller genome size, as was predicted. This may suggest that selfing does not result in genome reduction due to increased selection on selfish genetic elements including TEs. However, there are other factors affecting the dynamics of genome size in response to TEs which have not been controlled for within this analysis. For example, the patterns of TE content within a genome have been shown to be related to genome size itself (Novák *et al.*, 2020). Within plant species with a small or medium sized genome (up to around 10 Gbp), there is a positive relationship between repeat sequence content (of which TEs are a major component) and genome size with species with larger genomes having a higher proportion of repeat sequences within their genome. However, species with genomes larger than 10 Gbp show a negative relationship between repeat sequence and genome size, which Novák *et al.* (2020)suggest is due to the maintenance of non-repeat sequences derived from degraded repeat sequences within the genomes of species with larger genomes. Therefore, larger genomes may generally grow with time as they accumulate non-repeat sequences from selection against TEs. This might mean that any genome shrinkage due to increased selection against selfish genetic elements within selfing species may only be seen in species with smaller, more dynamic genomes.

Another important determinant of genome size in plants is ploidy. An organism’s ploidy level is the number of whole genomes each cell contains. Similar to humans, many plant species are diploid (contain two complete genomes in each cell; 2X). However, plants are also commonly polyploid (contain more than two complete genomes in each cell) and ploidy level can vary greatly between species up to 96-ploid (96X) within the fern *Ophioglossum reticulatum* (Leitch and Leitch, 2013). The comparison within pairs of selfing and outcrossing species suggests that ploidy may be an important factor in variation in genome size within UK species. Three quarters of the pairs showed at least a roughly 2-fold difference in genome size between the two species. Such a difference may be indicative of differences in ploidy between the species. Indeed, at least two pairs, *Veronica polita* and *V. chamaedrys* and *Senecio vulgaris* and *S. squalidus*, are known to show a two-fold difference in ploidy between the species (Albach and Grielhuber, 2005; Bennet and Smith, 1976). However, ratios of differences in genome size do not represent a very reliable method of determining whether a chromosome duplication event has occurred within a lineage. A better method of identifying whether the species within a pair differ in ploidy would be to compare the chromosome counts of the two species, with a roughly 2-fold difference in chromosome number being indicative of a genome duplication event having occurred in one species since their common ancestor. As genome duplication events are known to be associated with speciation within plants, it is not surprising that closely related pairs of species may differ in ploidy. However, plants of differing ploidy levels can show differing patterns of genome evolution (reviewed in Gerstein and Otto, 2009) so the 2X DNA content of species which differ in ploidy levels may not be comparable. Furthermore, ploidy may have a complex relationship with mating system as both selfing and perenniality, which is associated with outcrossing (Ma *et al.*, 2020), are thought to facilitate genome duplications (reviewed in Friedman, 2020; reviewed in Otto and Whitton, 2020).

Unfortunately, exclusion of pairs which differ in ploidy from the analysis would likely result in a sample which is too small to draw any meaningful conclusions about the general patterns of genome size between generally selfing and generally outcrossing species within the UK flora. Indeed, this analysis already suffered from a small sample size which likely limited its power. Use of gene sequence alignments to produce phylogenies may help overcome this problem of a small sample size by allowing resolution of the relationships between species within genera for which a literature review failed to resolve the relationships between the species of interest (*Poa*, *Ranunculus* and *Vicia*). This would allow identification of more pairs of closely related selfing and outcrossing species and increase the sample size.

The divergence time between selfing and outcrossing species within these pairs and the amount of evolution which has occurred since divergence will also likely differ between pairs. The application of gene sequence alignments could help control for this variation, as the production of phylogenies from gene sequence alignments allows estimation of branch lengths which could be used to control for differing levels of divergence between pairs of species.

As mating system is not responsible for a significant amount of variation within the UK flora, the increased selection on selfish genetic elements within selfing species is unlikely to be a major driver of genome size evolution within the UK flora. This evolution may instead be driven by other factors. For example, a species’ evolutionary history would likely be important as genome duplication events are important causes of genome size increase (reviewed in Pellicer *et al.*, 2018), although their effect is dependent on the length of time genome downsizing has had to occur following a duplication event (Wendel, 2015). Other factors affecting genome size can include ecological factors such as the availability of nitrogen and phosphorus, both required for the synthesis of DNA, to a plant (Guignard *et al.*, 2016).

## Phylogenetic Generalized Linear Models

The inclusion of more explanatory variables within initial pgls models, such as whether a species has a small or large genome or its ploidy level for the genome size analysis, would control for the variation in genome or range size they cause and reduce the amount of residual unexplained variation in the models. This would increase the power of the models and allow identification of smaller effects on genome or range size. As the factors affecting both genome and range size are numerous and are thought to interact in complex ways, a model including lots of them would likely soon become overly complex and model simplification would be required to identify the factors which had a significant effect on genome size.

As well as suffering from large amounts of unexplained variation, the pgls models for both the genome size and range size analyses were based on rather rudimentary phylogenies (see Appendix). These phylogenies could be improved by constructing new phylogenies from gene sequence alignments rather than relying on pre-existing literature. This would reduce the number of polytomies within the phylogenies and allow resolution of the relationships between species and genera which are currently uncertain (e.g., between and within *Juncus* and *Luzula*). Such phylogenies could also be constructed to include estimates of branch lengths, which would improve the pgls models by allowing the amount of independent evolution between species to be accounted for. The ongoing Darwin Tree of Life project (Darwin Tree of Life, 2020) would be invaluable for such an extension to this study as it would provide a source of high-quality genome, and therefore gene, sequences for any UK species of interest once complete.

## Range Size

Range size was found to be larger in outcrossing taxa in the complete flora dataset. Within subsets of the dataset, only aquatic and perennial plants showed a significant difference in range size between mating systems, with generally outcrossing taxa having a larger range size than generally selfing taxa. Mating system explained a significant amount of variation in range size within perennials but not within the complete native subset (although the interaction term between mating system and life history was significant in natives) or within aquatics.

The trend for outcrossing species to have larger average range sizes seems to be driven by perennial species, as it is the only subset investigated which showed the same trend as was found in the overall dataset and to have mating system explain a significant amount of variation in range size following phylogenetic correction. Although there was also a significant difference in range size between mating systems within aquatic species, mating system was not a significant factor in the pgls model which was fitted, suggesting that this difference may be driven by relatedness between species. For example, one family may have had a conserved outcrossing mating system and large range sizes.

The lack of difference in range size between mating systems within alpine and coastal species may be because the availability of suitable habitat is limited within the UK. The UK’s coastline is quite extensive but covers less area than inland areas and some habitats commonly associated with coastal species, such as sea cliffs or shingle beach, are more limited. It could therefore be this distinction between species limited to rarer coastal habitats and those which inhabit more widely distributed ones such as sand dunes which may instead determine range size of coastal species.

Alpine species within the UK often have a very limited range size, generally being confined to the Scottish Highlands, the Welsh mountains such as Snowdonia and the North Pennines. However, this suitable habitat is not distributed continuously across the UK, with different areas of highland acting more like “islands” of available habitat. This can be seen in the numbers of alpine species which are only found in one of these areas within the UK, such as *Minuartia stricta* which is confined to Upper Teesdale or *Cicerbita alpina* found only in the Cairngorms within the UK (BRC, 2020). Also, unlike many other species in the UK flora, some alpine species may have maintained populations within the British Isles during the most recent glaciation event so their range may be determined by the availability of suitable habitat during glaciation, such as exposed nunataks (Turrill, 1948), rather than by patterns of colonisation following the glaciation and modern-day habitat availability.

The lack of trend within non-native species may seem surprising given the proposed link between selfing and colonisation. Many non-natives are accidental introductions or garden escapes, so an increased ability to form naturalised populations and expand populations would be expected to be seen most strongly in these recently introduced species. The absence of a trend from this dataset could suggest that selfing species do not invade new areas more efficiently, and therefore may not have an advantage during colonisation of new areas. However, this may not represent the full picture. Non-native species will include those which have been purposefully introduced into the British Isles by humans such as trees used in forestry or ornamental plants within parkland, both of which may be more likely to be outcrossing taxa either due to being long-lived or due to a human preference for large, showy flowers associated with pollinator attraction. It also includes arable weeds, which were inadvertently introduced to large areas by humans during the introduction of crops. Therefore, plants which are introduced widely within the UK by human intervention may have larger range sizes than those which are invading from isolated introduction or garden escape events, so an investigation into the origins of these species may be useful to distinguish between directly anthropogenically inflated range sizes and range sizes determined by colonisation and range expansion ability.

## Range Size and Life History

Range size was associated with mating system only within perennial native plants and not annuals. The interaction between life history and mating system, identified as significant within the pgls model for the native species, is well established. Longer-lived plants are less likely to have a self-fertilizing mating system, possibly due to a higher number of mitosis events increasing the mutational load in longer-lived species (Scofield and Schultz, 2006) and experiencing additional costs of inbreeding depression due to impacts on the ability of an individual to survive between breeding events and therefore increase their lifetime fecundity (Morgan *et al.*, 1997). This is reflected in the proportion of selfing species being lower in perennials than annuals.

Perennials are a lot more diverse than annuals in terms of some life history traits. For example, perennials can vary from biennials which live for two years to species such as the Great Basin bristlecone pine (*Pinus longaeva*) which can live for thousands (Lanner and Connor, 2001). The timing of reproduction can also vary. Some flower and reproduce within their first year whilst others have non-reproductive juvenile phases which can last years. These factors may also be expected to covary with mating system due to the increased costs of inbreeding depression in longer-lived species. However, they were not accounted for within this study, so the possibility that the significant effect of mating system on range size is due to an association between outcrossing and other life history traits such as a longer lifespan cannot be ruled out. To further this study and confirm that the observed trend is truly due to differences between mating systems and not due to such traits which may covary with mating system, a more complex model which includes more life history factors such as longevity and the length of the juvenile phase could be produced.

## Potential Confounding Factors in Range Size

Mating system is also known to be associated with life form, with herbaceous plants being more likely to self-fertilize than woody ones (Barrett *et al.*, 1996). Life form explained significant variation in range size in pgls models for both native and perennial native species, suggesting that it is a major contributor to range size differences within the British Isles. There was also an overlap between the variation range size described by mating system and life form, as would be expected between covarying traits, with mating system ceasing to describe significant variation on its own once shared variation with life form and life history was accounted for in the native subset. However, within perennials, mating system is a significant factor in predicting range size independent of life form. This shows that, although the difference in median range size ceases to be significant following the removal of phanerophytes, the difference in range size between mating systems is not purely due to self-fertilization being less common in life forms such as phanerophytes which tend to have larger range sizes.

The long history of land use changes, such as forest clearance starting in the early Neolithic (Turrill, 1948), and land management, such as the grazing of meadowland by livestock or the exclusion of trees from moorland, may be expected to result in unusual patterns of regional range size within plants where plants which have a small global distribution may have a large range size within the British Isles or vice versa. The strong correlation between UK range size and the number of continents a species is present on would suggest that this is not an important factor within either the native or perennial subsets. However, continental presence counts are a poor proxy for global distribution. For example, species confined to the Mediterranean Basin are commonly present on three continents (Europe, Asia and Africa) while those with a near ubiquitous spread across the whole of Eurasia will only be present on two. To prevent this bias, counts of the number of biogeographic realms or regions a species is present in globally may have been more appropriate as they are more reflective of variation in climate and boundaries to dispersal than the continents a species is present on. However, this alternative method would still suffer from the fact that large regions of some continents are poorly surveyed within the GBIF database. The disparity in surveying between regions means that the data is unlikely to reflect the true global distribution of species, with species being more likely to be falsely determined as absent from continents with larger areas which are poorly surveyed such as Africa and South America.

## Theoretical Explanations for a Smaller Range Size in Selfing Species

The negative association between selfing and range size agrees with the ecological niche hypothesis that selfing species should have a smaller range size due to having smaller niches because of lower genetic diversity. However, studies into the relationship between niche size and range size within the UK (Thompson and Ceriani, 2003) and Germany (Kolb *et al.*, 2006) suggest that niche size is not an important factor in determining range size. There has been the suggestion that niche breadth may be more important within certain life history groups (e.g., Brändle *et al.*, 2003), but it is hypothesised to be more important in annuals. This therefore suggests that interactions between mating system and niche breadth within perennials is less likely to determine differences in range size.

Alternatively, metapopulation dynamics may cause the observed relationship between range size and mating system. Although many predictions of the metapopulation hypothesis for range size regarding mating system predict that selfing species should have a larger range size, some suggest selfing species may indeed have a smaller range size. For example, Massol and Cheptou’s (2011)expansion of Baker’s Law suggest that selfing species should only have a colonisation advantage, and therefore would only be expected to be able to establish larger ranges due to this advantage, when dispersal to a newly colonised site is purely one-sided. This would be the case for many oceanic islands, but not for non-oceanic islands so close to the continental mainland as the British Isles. The British Isles’ status as islands has not even been maintained throughout their history, with land bridges connecting Great Britain to both Europe and Ireland during and after the last glaciation (Turrill, 1948). This peninsular Britain would therefore have experienced colonisation where there was no barrier to dispersal or gene flow out of Britain and back to the modern-day continental mainland, under which conditions selfing is not predicted to be linked to dispersal (Massol and Cheptou, 2011). Even following the inundation of Doggerland and the formation of the North Sea splitting Britain from mainland Europe, there would likely still be dispersal in both directions even if it became more asymmetric, under which conditions selfing species may still not have a colonisation advantage (Massol and Cheptou, 2011). This suggests that selfing species would not have had an increased ability to colonise the British Isles following the last glacial maximum which may have otherwise allowed them to achieve large range sizes within the British Isles.

Another factor which may be important to range size under the metapopulation hypothesis is the ability of a plant to disperse. If a species does not disperse from its current site, it will be unable to colonise new areas and increase its range size. Under conditions where there is variability in the efficacy of pollination, such as where population density is patchy or where there is temporal variation in the availability of pollinators, Massol and Cheptou (2010) predict that one of two strategies should be selected for. Either species will be selected to become outcrossing and have an increased capacity for dispersal or it will be selected to self-fertilize, in which case dispersal will be limited. Therefore, if there is variability in pollination within the British Isles either in the modern day or historically, it may have selected for outcrossing species with increased dispersal which went on to increase their range size and selfing species who did not experience an increase in dispersal capability and therefore retained relatively static range sizes.

## Relationships with Range versus Genome Size

Range size was found to be significantly correlated with mating system in at least some subsets of the UK flora, whilst genome size was not significantly correlated. This suggests that extrinsic factors may have a stronger association with mating system than intrinsic factors within the UK flora. This may be because extrinsic factors, such as ecological pressures, are more likely to impose selective pressures on a species to evolve a particular mating system whilst intrinsic ones are more likely to evolve in response to the transition and therefore may respond differently or to varying degrees to the evolution of selfing. Intrinsic factors are also influenced by the evolutionary history of a species, for example the genome size of a species will be determined by historic genome expansions or gene duplication events as well as any current evolutionary factors which may be selecting for genome expansion or reduction. However, the impact of mating system on range size was also modulated by other factors, such as life history and life form traits, with an effect of mating system on range size only being found in perennials in which life form and mating system explain shared variation in range size. This highlights a fundamental point of biology, the impact of any one factor, such as mating system, will be likely to interact with and be modulated by a vast array of other factors in a complex manner which is difficult to disentangle.

## Implications for Theories Involving Selfing

The lack of a significant relationship between genome size and mating system within the UK flora suggests that selfing is not an important factor in the determination of a species’ genome size and the trends observed at a familial or generic level, such as those found by Albach and Grielhuber (2004), may not be generalizable to complete flora. However, there was also no significant relationship within the Poaceae, Asteraceae, Fabaceae, Brassicaceae, Ranunculaceae, Plantaginaceae or Rosaceae families. As a relationship between selfing and genome size had been observed in previous studies of different families, it may be that none of the families analysed independently within this study show strong genome shrinkage in response to selfing. This could be linked to the active TE load of a family, as genome size reduction in selfing has been hypothesised to be related to stronger selection against selfish genetic elements. Alternatively, the lack of control for variation in genome size driven by other traits may have meant the analyses were not strong enough to detect any variation driven by mating system which may be attributable to selfing.

As selfing was not universally associated with any difference in range size across the UK flora, the effect of mating system on range size appears to be modulated by both habitat (in the case of alpine and coastal species) and life history. However, selfing perennial species having a smaller range size on average suggests that, in the UK at least, characteristics which would be associated with increased range size such as increased dispersal or ability to colonise new areas are unlikely to be associated with selfing within perennials. Therefore, selfing must have other advantages for the species that adopt it, for instance reproductive assurance or the maintenance of advantageous genetic backgrounds across generations.

# ACKNOWLEDGEMENTS

Drs Matthew Hartfield and Catherine Kidner acted as supervisors for this project, and provided invaluable guidance on ideas, relevant literature and the direction of the project throughout. Sean O’Brien compiled the UK Flora Database and cleaned some vital entries, such as range size and mating system, prior to project as described in the first section of the “Methods and Materials”. Dr Richard Milne provided guidance on relevant literature concerning the history of the UK flora.

# REFERENCES

Ågren, J. A., Wang, W., Koenig, D., Neuffer, B., Weigel, D., and Wright, S. I. (2014). Mating system shifts and transposable element evolution in the plant genus Capsella. *BMC Genomics*, 15: 602; DIO: <https://doi.org/10.1186/1471-2164-15-602>

Albach, D. C., and Greilhuber, J. (2004). Genome Size Variation and Evolution in Veronica. *Annals of Botany*, 94: 897-911.

Asmussen, C. B., and Liston, A. (1998). Chloroplast DNA characters, phlogeny, and classification of *Lathyrus* (Fabaceae). *American Journal of Botany*, 85 (3): 387-401.

Barrett, S. C. H., Harder, L. D., and Worley, A. C. (1996). The comparative biology of pollination and mating in flowering plants. *Phil. Trans. R. Soc. Lond. B*, 351; 1271-1280.

Bennet, M. D., and Smith, J. D. (1975). Nuclear DNA amounts in angiosperms. *Phil. Trans. R. Soc. Lond. B*, 274: 227-274.

Brändle, M., Stadler, J., Klotz, S., and Brandl, R. (2003). Distributional range size of weedy plant species is correlated to germination patterns. *Ecology*, 84 (1): 136-144.

BRC (2020). *Online Atlas of the British and Irish Flora*. [online] Available at: <https://www.brc.ac.uk/plantatlas/> (Accessed 14 April 2021)

Brown, J. H. (1984). On the relationship between abundance and distribution of species. *The American Naturalist*, 124 (2): 255-279.

Burt, A., and Trivers, R. (1998). Selfing DNA and breeding system in flowering plants. *Proc. R. Soc. Lond. B*, 265: 141-146.

Byng, J. W., Smets, E. F., van Vugt, R., Bidault, E., Davidson, C., Kenicer, G., Chase, M. W., and Christenhusz, M. J. M. (2018). *The phylogeny of angiosperms poster: a visual summary of APG IV family relationships and floral diversity*. Plant Gateway Ltd. Available at: <http://www.plantgateway.com/poster/>.

Carlsen, T., Bleeker, W., Hurka, H., Elven, R., and Brochmann, C. (2009). Biogeography and Phylogeny of *Cardamine* (Brassicaceae). *Annals of the Missouri Botanical Garden*, 96 (2): 215-236.

Carolan, J. C., Hook, I. L. I., Chase, M. W., Kadereit, J. W., and Hodkinson, T. R. (2006). Phylogenetics of *Papaver* and Related Genera Based on DNA Sequences from ITS Nuclear Ribosomal DNA and Plastid *trnL* Intron and *trnL-F* Intergenic Spacers. *Annals of Botany*, 98 (1): 141-155.

Charlesworth, B., and Langley, C.H. (1986). The evolution of self-regulated transposition of transposable elements. *Genetics*, 112: 359-383.

Charlesworth, D., and Willis, J. H. (2009). The genetic of inbreeding depression. *Nature Reviews Genetics*, 10: 783-796.

Chester, M., Cowan, R. S., Fay, M. F., and Rich, T. C. G. (2007). Parentage of endemic *Sorbus* L. (Rosaceae) species in the British Isles: evidence from plastid DNA. *Botanical Journal of the Linnean Society,* 154 (3): 291-304.

Darwin Tree of Life (2020). *The Darwin Tree of Life*. [online] Available at: <https://www.darwintreeoflife.org/> (Accessed 21 April 2021)

Drábková, L., Krischner, J., and Vlček, C. (2006). Phylogenetic relationships within *Luzula* DC. and *Juncus* L. (Juncaceae): A comparison of phylogenetic signals of *trn*L-*trn*F intergenic spacer, *trn*L intron and *rbc*L plastome sequence data. *Cladistics*, 22: 132-143.

Eckardt, N. A. (2011). A Sense of Self: Exploring the Selfing Syndrome in *Capsella*. *The Plant Cell*, 23 (9) 3086; DIO: <https://doi.org/10.1105/tpc.111.230913>

Eklund, A. (2016). *beeswarm: The Bee Swarm Plot, an Alternative to Stripchart*. Available at: <https://CRAN.R-project.org/package=beeswarm>.

Ellison, N. W., Liston, A., Steiner, J. J., Williams, W. M., and Taylor, N. L. (2006). Molecular phylogenetics of the clover genus (*Trifolium*-Leguminosae). *Molecular Phylogenetics and Evolution*, 39 (3): 688-705.

Friedman, J. (2020). The Evolution of Annual and perennial Plant Life Histories: Ecological Correlates and Genetic Mechanisms. *Annual Review of Ecology, Evolution, and Systematics*, 51: 461-481.

Fitter, A. H., and Peat, H. J. (1994). The Ecological Flora Database. *Journal of Ecology*, 82: 415-425.

Fox, J., and Weisberg, S. (2019). *An R Companion to Applied Regression*. Third Edition. Thousand Oaks CA: Sage Publications.

GBIF.org (2021). *GBIF Home* Page. [online] Available at: <https://www.gbif.org> (Accessed 14 April 2021)

Gerstein, A. C., and Otto, S. P. (2009). Ploidy and the Causes of Genomic Evolution. *Journal of Heredity*, 100 (5): 571-581.

Goodwillie, C., Kalisz, S., and Eckert, C. G. (2005). The Evolutionary Enigma of Mixed Mating Systems in Plants: Occurrence, Theoretical Explanations, and Empirical Evidence. *Annu. Re. Ecol. Evol. Syst.*, 36: 47-79.

Grossenbacher, D. L., Brandvain, Y., Auld, J. R., Burd, M., Cheptou, P.-O., Conner, J. K., Grant, A. G., Hovick, S. M., Pannell, J. R., Pauw, A., Petanidou, T., Randle, A. M., de Casas, R. R., Vamosi, J., Winn, A., Igic, B., Busch, J. W., Kalisz, S., and Goldberg, E. E. (2017). Self-compatibility is over-represented on islands. *New Phytologist*, 215: 469-478.

Grossenbacher, D., Runquist, R. B., Goldberg, E. E., and Brandvain, Y. (2015). Geographic range size is predicted by plant mating system. *Ecology Letters*, 18 (7): 706-713.

Guignard, M. S., Nichols, R. A., Knell, R. J., Macdonald, A., Romila, C.-A., Trimmer, M., Leitch, I. J., and Leitch, A. R. (2016). Genome size and ploidy influence angiosperm species’ biomass under nitrogen and phosphorus limitations. *New Phytologist*, 210: 1195-1206.

Hamrick, J. L., and Godt, M. J. W. (1996). Effects of life history traits on genetic diversity in plant species. *Phil. Trans. R. Soc. Lond. B*, 351: 1291-1298.

Hanski, I. (1998). Metapopulation dynamics. *Nature*, 396: 41-49.

Hill, M. O., Preston, C. D., and Roy, D. B. (2004). *PLANTATT - Attributes of British and Irish Plants - Spreadsheet*. Huntingdon: Centre for Ecology and Hydrology.

Jarne, P., and Auld, J. R. (2006) Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. *Evolution*, 60 (9): 1816-1824.

Kirkpatrick, M., and Barton, N. H. (1997). Evolution of a Species’ Range. *The American Naturalist*, 150 (1): 1-23.

Kolb, A., Barsch, F., and Diekmann, M. (2006). Determinants of local abundance and range size in forest vascular plants. *Global Ecology and Biogeography*, 15: 237-247.

Lanner, R. M., and Connor, K. F. (2001). Does bristlecone pine senesce? *Experimental Gerontology*, 36: 675-685.

Lee, S.-I. and Kim, N.-S. (2014). Transposable Elements and Genome Size Variations in Plants. *Genomics Inform*, 12 (3): 87-97.

Leitch, I. J., Johnston, E., Pellicer, J., Hidalgo, O., and Bennett, M. D. (2019). *Plant DNA C-values Database*. [online] Available at: <https://cvalues.science.kew.org/>. (Accessed: 14 April 2021).

Leitch, I. J., and Leitch, A. R. (2013). Genome Size Diversity and Evolution in Land Plants. In: Grielhuber, J., Dolezel, J., and Wendel, J. (eds) *Plant Genome Diversity Volume 2*. Vienna: Springer Vienna, pp. 307-322.

Ma, Y., Barrett, S. C. H., Wang, F.-Y., Deng, J.-C., and Bai, W.-N. (2020). Do annual and perennial populations of an insect-pollinated plant species differ in mating system? *Annals of Botany,* mcaa178; DIO: <https://doi.org/10.1093/aob/mcaa178>.

Massol, F., and Cheptou, P.-O. (2010). Evolutionary syndromes linking dispersal and mating system: the effect of autocorrelation in pollination conditions. *Evolution*, 65 (2): 591-598.

Massol, F., and Cheptou, P.-O. (2011). When should we the evolutionary association of self-fertilization and dispersal? *Evolution*, 65 (5): 1217-1220.

Morgan, M. T. (2001). Transposable element number in mixed mating populations. *Genet. Res., Cam.*, 77: 261-275.

Morgan, M. T., Schoen, D. J., and Bataillon, T. M. (1997). The evolution of self-fertilization in perennials. *The American Naturalist*, 150: 618-638.

Morris, J., *et al.* (2016). *Biology: How Life Works*. Second Edition. Macmillan Education.

Moscone, E. A., Samuel, R., Schwarzacher, T., Shweizer, D., and Pedrosa-Harand, A. (2007). Complex rearrangements are involved in *Cephalanthera* (Orchidaceae) chromosome evolution. *Chromosome Research* 15: 931-943.

Novák, P., Guignard, M. S., Neumann, P., Kelly, L. J., Mlinarec, J., Koblížkova, A., Dodsworth, S., Kovařík, A., Pellicer, J., Wang, W., Macas, J., Leitch, I. J., and Leitch, A. R. (2020). Repeat-sequence turnover shifts fundamentally in species with large genomes. *Nature Plants*, 6: 1325-1329.

Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., and Pearse, W. (2018). *caper: Comparative Analyses of Phylogenetics and Evolution in R.* Available at: <https://CRAN.R-project.org/package=caper>.

Otto, S. P. (2009). The Evolutionary Enigma of Sex. *The American Naturalist*, 174: Supplement; DIO: <https://doi.org/10.1086/599084>.

Otto, S. P., and Whitton, J. (2020). Polyploid incidence and evolution. *Annual Review of Genetics*, 34: 401-437.

Paradis, E., and Schliep, K. (2019). Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35: 526-528.

Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I. J. (2018). Genome Size Diversity and Its Impact on the Evolution of Land Plants. *Genes*, 9 (2): 88; DIO: <https://doi.org/10.3390/genes9020088>.

Pelser, P. B., Gravendeel, B., and van der Meijden, R. (2002). Tackling speciose genera: species composition and phylogenetic position of *Senecio* sect. *Jacobaea* (Asteraceae) based on plastid and nrDNA sequences. *American Journal of Botany*, 89 (6): 929-939.

POWO (2019). *Plants of the World Online*. [online] Available at: <http://www.plantsoftheworldonline.org/>. (Accessed: 14 April 2021).

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Available at: <http://www.R-project.org/>.

Randle, A. M., Slyder, J. B., and Kalisz, S. (2009). Can differences in autonomous selfing ability explain differences in range size among sister-taxa pairs of Collinsia (Plantaginaceae)? An extension of Baker’s Law. *New Phytologist*, 183: 618-629.

Robertson, A., Newton, A. C., and Ennos, R. A. (2003). Multiple hybrid origins, genetic diversity and population genetic structure of two endemic *Sorbus taxa* on the Isle of Arran, Scotland. *Molecular Ecology*, 13: 123-134.

Roessler, K., Muyle, A., Diez, C. M., Gaut, G. R. J., Bousios, A., Stilzer, M. C., Seymour, D. K., Doebley, J. F., Liu, Q., and Gaut, B. S. (2019). The genome-wide dynamics of purging during selfing in maize. *Nature Plants*, 5: 980-990.

RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PCS, Boston, MA. Available at: <https://www.rstudio.com>.

Ruhsam, M., Squirrell, J., Gornall, R.J., French, G. C., Pullan, M., and Hollingsworth, P. M. (2020). *Genetic Flora of the British Isles database.* [online] Available at: <https://websites.rbge.org.uk/geneticflora/gflora.php> (Accessed 14 April 2021).

Scofield, D. G., and Schultz, S. T. (2006). Mitosis, stature and evolution of plant mating systems: low-Φ and high-Φ plants. *Proc. R. Soc. B*, 273: 275-282.

Sincard, A., and Lenhard, M. (2011). The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Annals of Botany*, 107: 1433-1443.

Stebbins, G. L. (1957). Self fertilization and population variability in the higher plants. *The American Naturalist*, 91 (861): 337-354.

Taskova, R. M., Albach, D. C., and Grayer, R. J. (2004). Phylogeny of *Veronica* – a Combination of Molecular and Chemical Evidence. *Plant Biol.*, 6: 673-682.

Thompson, K., and Ceriani, R. M. (2003). No relationship between range size and germination niche width in the UK herbaceous flora. *Functional Ecology*, 17: 335-339.

Turrill, W. B. (1948). *British Plant Life.* First Edition. London: Collins, pp. 35-48.

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth Edition. New York: Springer.

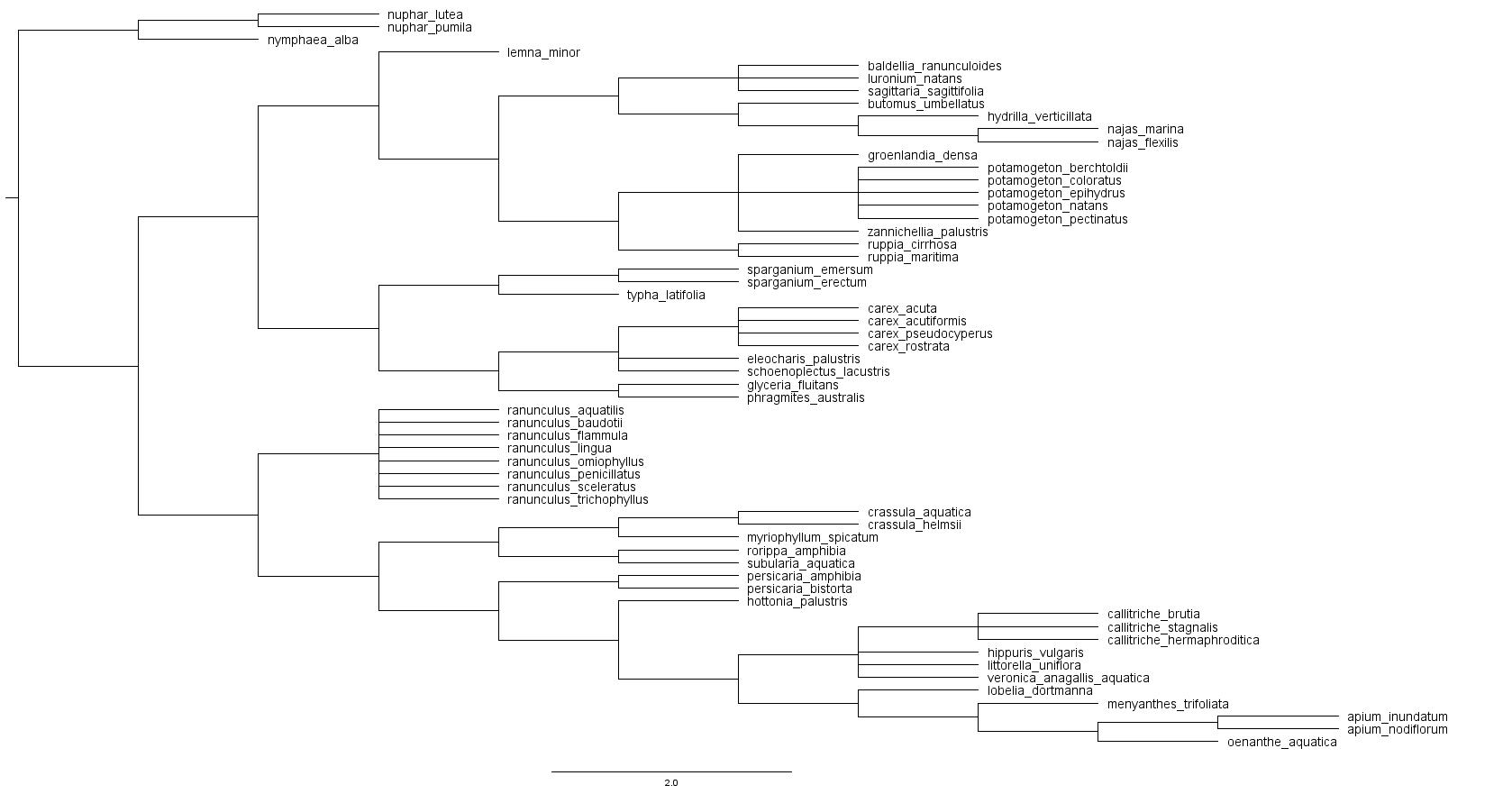
Wendel, J. F. (2005). The wondrous cycles of polyploidy in plants. *American Journal of Botany*, 102 (11): 1753-1756.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* New York. Springer-Verlag New York.

Wright, S. I., Ness, R. W., Foxe, J. P., and Barrett, S. C. (2008). Genomic consequences of outcrossing and selfing in plants. *Int. J. Plant Sci.*, 169 (1): 105-11

**APPENIDIX**

Appendix Fig. 1- A phylogeny of species present in the UK Flora Database aquatic range subset as used to produce phylogenetic generalized linear models.



Diagram, schematic

Description automatically generated

a)

Diagram, schematic

Description automatically generated

b)

Diagram, schematic

Description automatically generated

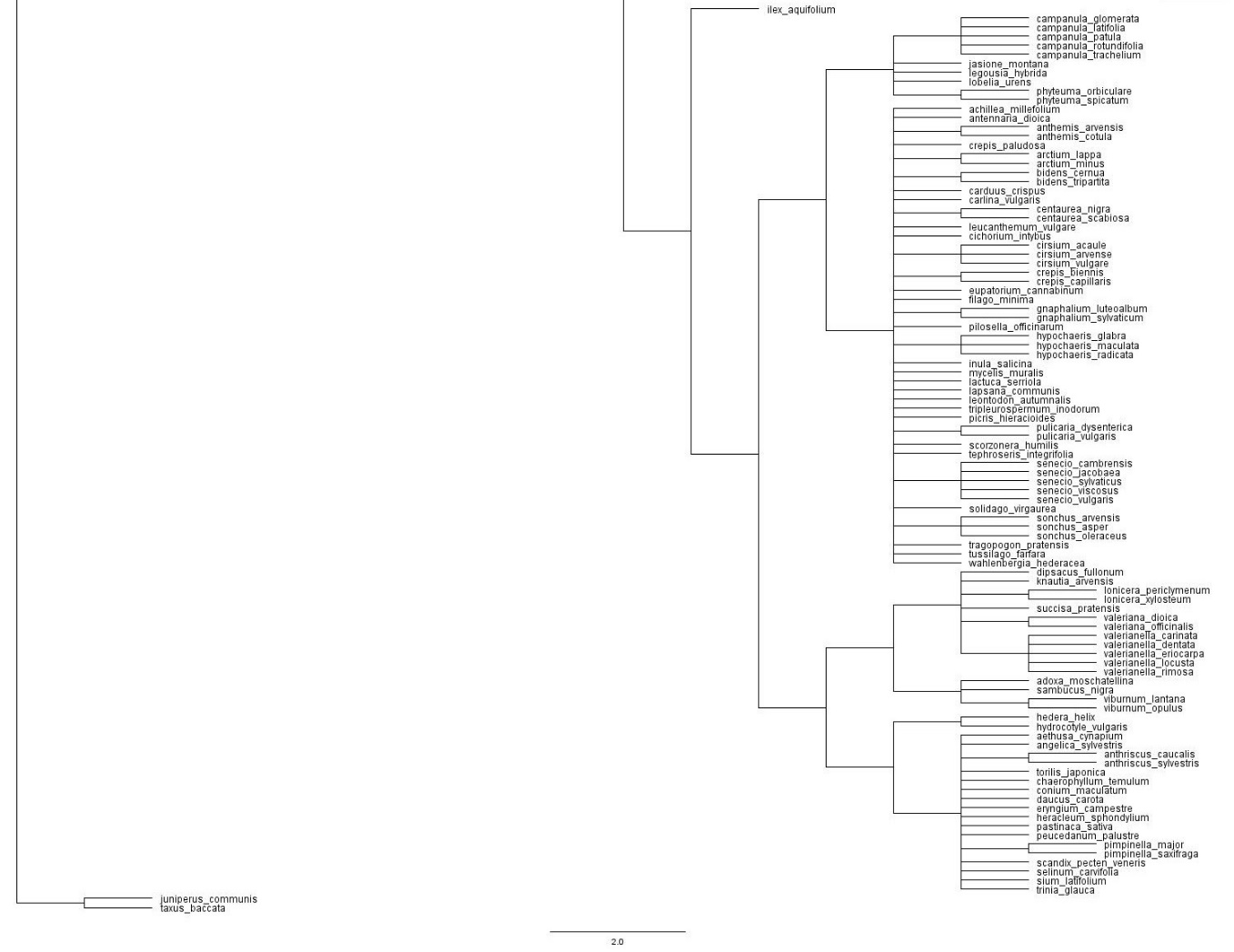
c)

Diagram, schematic

Description automatically generated

d)

e)



Appendix Fig. 2 a-e- Sequential segments of a phylogeny of species present in the UK Flora Database native range subset as used to produce phylogenetic generalized linear models.

Diagram, schematic

Description automatically generated

a)

A picture containing text, antenna

Description automatically generated

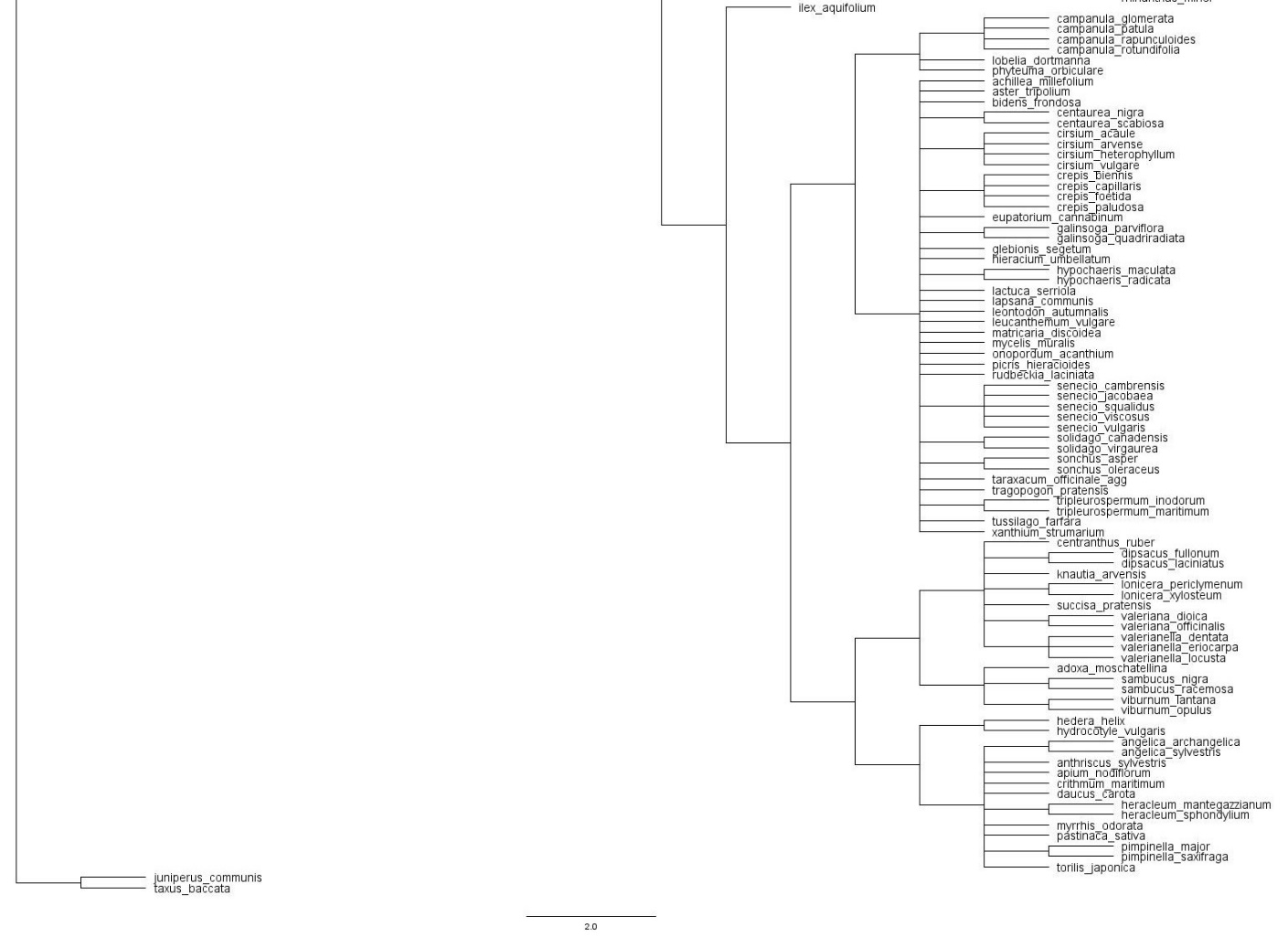
b)

Diagram, schematic

Description automatically generated

c)

d)



Appendix Fig. 3 a-d- Sequential segments of a phylogeny of species present in the UK Flora Database genome dataset as used to produce phylogenetic generalized linear models.