**Analysis of the association between selfing and exemplar extrinsic and intrinsic traits in UK flora**

**Introduction**

Self-fertilization, alternatively known as “selfing”, is when the gametes involved in sexual reproduction come from the same individual. It is seen to varying degrees in both animals (Jarne and Auld, 2006) and plants (Goodwillie *et al.*, 2005). However, the existence of such a fertilization strategy is surprising given that many organisms avoid inbreeding, of which selfing is the most extreme form. This avoidance may occur due to inbreeding depression, the reduction in fitness which occurs when closely related organisms breed. Inbreeding depression is thought to result from increased homozygosity in offspring, which can both unmask recessive deleterious alleles and lead to a loss of heterozygous advantage effects (reviewed in Charlesworth and Willis, 2009). One theory as to why selfing may occur despite this disadvantage is that some organisms, such as sessile organisms or those with low population densities, may benefit from the reproductive assurance of selfing (Morris *et al.*, 2016). Alternatively, organisms living in stable environments may not benefit as much from the benefits of outcrossing (such as the increase in the genetic diversity of the offspring and the production of new combinations of alleles) as they would likely already be adapted to their environment due to the action of natural selection. The increased transmission of their genes resulting from contributing both gametes to offspring may therefore be more advantageous, with the shuffling of genetic backgrounds which results from outcrossing possibly acting deleteriously by breaking up adaptive genetic backgrounds. Also, although inbreeding depression may be detrimental in the short term, the exposure of deleterious recessive alleles to selection will allow their purging (Wright *et al.*, 2008), possibly decreasing the mutational load of a selfing population in the long term, meaning long-term selfing populations should have a low recessive mutation load and therefore have very low levels of inbreeding depression.

Selfing organisms often possess similar characteristics. Some of these, such as a large range size, are likely shared as they drive the evolution of selfing. Others, such as loss of self-incompatibility, may be required for the transition to selfing. Finally, some shared characteristics could result from the evolution of selfing, including some elements of so called “selfing syndrome”, such as a reduction or loss of mechanisms to attract a mate or pollinator and a change in the relative investment in male and female function (Sincard and Lenhard, 2011).

Range size is a trait known to be associated with selfing, with a positive relationship between range size and selfing being shown by Grossenbacher *et al*’s (2015) metanalysis of the range size of a large sample of plants whose combined native ranges included every continent except Antarctica. This relationship has also been demonstrated on a generic level, with *Collinsia* species which are more likely to self-fertilise being shown to have larger ranges (Randle *et al.*, 2009). Selfing is thought to enable larger range sizes as it should theoretically increase the ability of individuals to colonise new areas as they aren’t constrained by the need to find a mate. The link between selfing and colonization is exemplified by Baker’s Law, that islands have a higher proportion of selfing species (Stebbins, 1957), which is supported by studies such as Grossenbacher *et al.*’s(2017) study into the frequency of self-incompatibility (as a proxy for non-selfing) in continental and island plant populations that found a higher proportion of island plants are self-compatible and therefore more likely to be able to self-fertilize. Furthermore, selfing may also allow rapid population growth and limit the impact of Allee effects on a founding population. If a population is selfing prior to colonisation, it may have previously experienced purging of deleterious recessive alleles and therefore have a low mutational load, enabling it to avoid any reduction in fitness associated with the expression of deleterious recessive alleles due to the initially small gene pool of founding populations. Alternatively, the small population sizes of founding populations may promote the evolution of or an increase in self-fertilization for the reasons above (although, without previous purging, these populations would still likely initially experience inbreeding depression because of deleterious recessive alleles).

Selfing may also be associated with a reduction in genome size. An association has been shown by comparative studies, such as that by Albach and Greilhuber (2004)which found that selfing taxa within the plant family Veroniceae tend to have smaller genomes. A reduction in genome size due to selfing has also been observed experimentally in maize lines by Roessler *et al.* (2019), who self-fertilised 11 outbred maize lines to produce homozygous lines which went through 6 generations of self-fertilisation. However, out of the 11 experimentally inbred lines, only 3 showed a reduction in genome size. Changes in genome size in the lines were found to be most strongly correlated with a reduction in transposable elements. This supports previous theoretical work which suggest that selection on selfish genetic elements such as transposable elements (Morgan, 2001;Charlesworth and Langley, 1986) or B chromosomes (Burt and Trivers, 1998) may be stronger in selfing populations. As such elements are a known driver of genome expansion (Lee and Kim, 2014), selection against them may be involved in genome size reduction, as is suggested by the Roessler *et al.* (2019)study. However, the relationship between fertilisation strategy and transposable element content has failed to be upheld in some studies, such as Ågren *et al.*’s (2014) comparative study in selfing and self-incompatible species within the *Capsella* genus which found that the number of transposable elements wasn’t higher in the self-incompatible *C. grandiflora* than a related selfing species, *C. rubella* (although it was when *C.* grandiflora was compared to a different selfing species, *C. ortientalis*).

Understanding the relationships between selfing and other factors, both extrinsic (such as range size or ecological function) and intrinsic (such as morphology or genome size and content), is important for the understanding of why selfing is sometimes favourable despite its apparent detrimental effects, and how such detrimental effects can be limited or avoided. Furthermore, the relationship between genome size and selfing is not yet fully understood. Whilst a reduction in genome size may be associated with selfing, the cause of such a reduction remains unresolved, although as shown above some research suggests a link with purging of transposable elements. Investigating genome size will allow investigation of whether it is correlated with selfing across all taxa in a sample and therefore may be directly influenced by selfing. Alternatively, if the association is only seen in some families with shared traits (such as high levels of active transposable elements), the association of selfing and genome size may be a secondary relationship dependant on the effect of selfing on such a trait. Range size, being a well-established covariant with selfing, will provide a good characteristic for comparison with genome size to establish the strength of any association between genome size and selfing.

The UK Flora Database (Fitter and Peat, 1994) provides a good dataset for researching the relationship between selfing and other ecological characteristics as it is based on a well-studied and understood floral assembly. Although there are few genetic characteristics included in the dataset, it does include information on genome size and other datasets, or literature reviews, could be used to gather information on other genetic characteristics if required. The dataset isn’t as geographically large as some studies, such as that conducted by Grossenbacher *et al* (2015), but it does contain information on a taxonomically diverse group of plants which will allow the conformation or otherwise of trends previously only observed in specific genera or families. As Britain has been an island for much of its recent history since the end of the last glacial maximum (following which much of the UK flora has been assembled), selfing species may be expected to be overrepresented in it due to Baker’s Law, meaning that it may provide a strong dataset for investigating relationships between selfing and other traits. Furthermore, previous work by Sean O’Brien means that some of the data in the dataset has already been pre-cleaned, reducing the amount of data cleaning required.

This study will aim to investigate the strength of the association between selfing and exemplar extrinsic and intrinsic effects of a selfing lifestyle: the expansion of range size and the reduction of genome size.

**Methods**

To allow analysis of 2C DNA content from the UK Flora Database, the data will require cleaning up as some entries currently contain multiple values. 2C DNA content will be cleaned with reference to the Plant DNA C-values Database (Leitch *et al*, 2019) to ensure there aren’t multiple entries for individual species. This won’t be required for fertilisation method or range size data as it has already been pre-cleaned by Sean O’Brien.

The normality of the 2C DNA content and range size data will be tested using a Shapiro Wilk test. If the data isn’t normally distributed, transformations such as logarithmic or inversion transformations will be attempted to produce a normal distribution. Depending on whether the data has successfully been transformed into a normal distribution, a one-way ANOVA or a Kruskal-Wallis test will be used to test whether the variables’ values differ significantly between fertilisation modes (see Fig. 1).

As an initial test for phylogenetic dependence, this analysis will be repeated in families which have adequately large sample sizes (identified via a power analysis) and both selfing and non-selfing taxa. As only 24% of species in the database currently have family information, and only 4 families currently have species assigned to them, data on the families which species belong to will be added to the database with reference to the Plants of the World online database (POWO, 2019). If the trend observed across the families is relatively uniform and similar to that observed following analysis of the complete sample, it can be inferred that phylogeny has a small impact on the relationship. This will be confirmed by carrying out phylogenetically independent comparisons at the familial level. A Spearman’s rank correlation coefficient will be used to determine whether selfing varies significantly with 2C DNA content and range size, respectively, between phylogenetically independent comparisons, following the conversion of fertilization mode data to ranked data ranked from least to most selfing.

To allow comparison of the ability of 2C DNA content and range size, respectively, to predict selfing in plants, a linear model with fertilization mode (transformed into a numerical dataset) as the response variable and either 2C DNA content or range size as the explanatory variable will be produced. To allow for this, the categorical selfing data with the most categories will be used to produce numerical selfing data with 5 bins, 0 (no selfing) to 4 (complete selfing). If the phylogenetically independent contrast shows that these variables aren’t independent of phylogeny, these linear models could be produced for individual families rather than the whole dataset. The confidence in these models and the strength of the linear correlation between fertilisation mode and the respective explanatory variables will then be used to determine how well each explanatory variable explains variation in fertilisation mode.

**Timetable**

The combined allocation of time for this research project, as well as associated training and the write-up of the resulting thesis, will be 503 hours over the course of 15 weeks from the 11th of January 2021 (see Fig. 2).

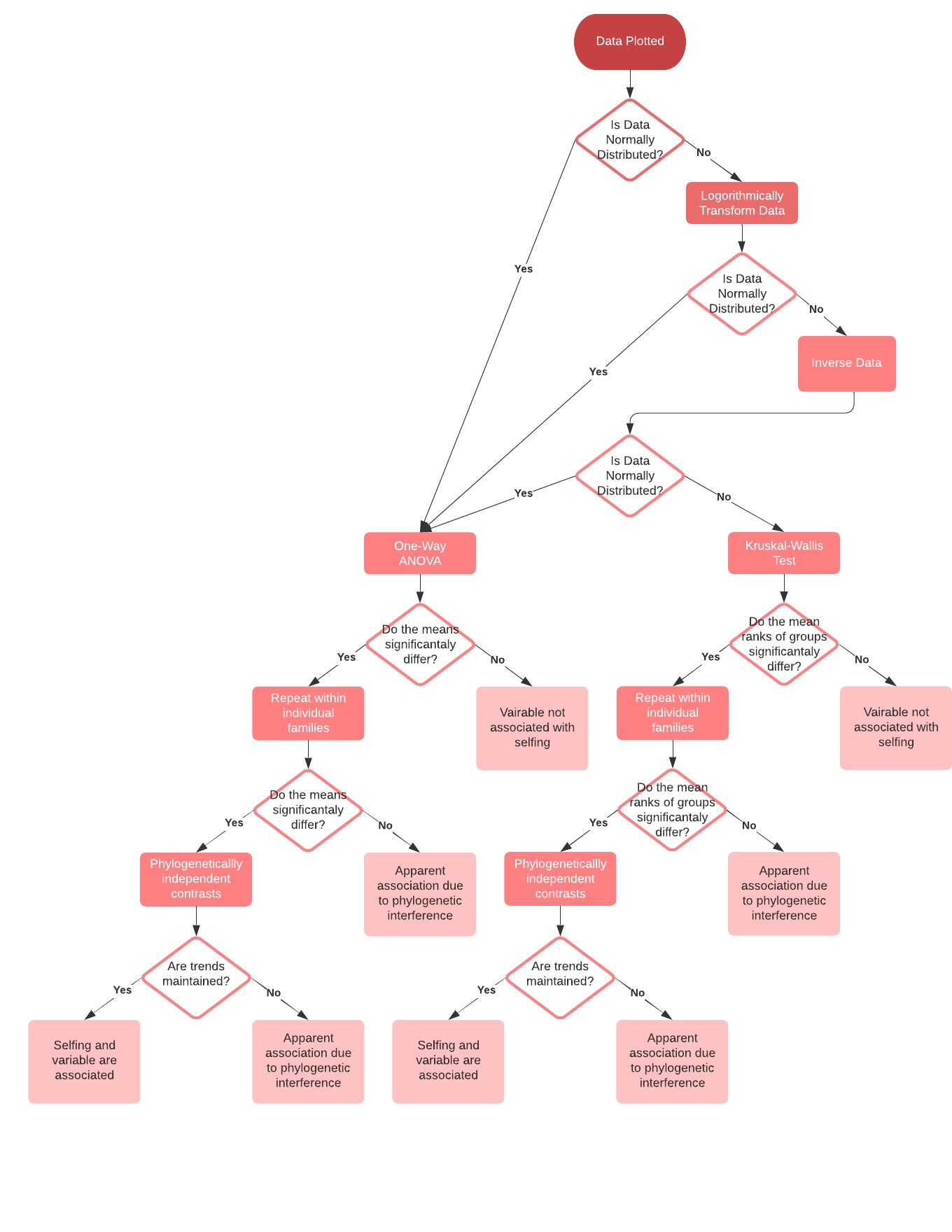
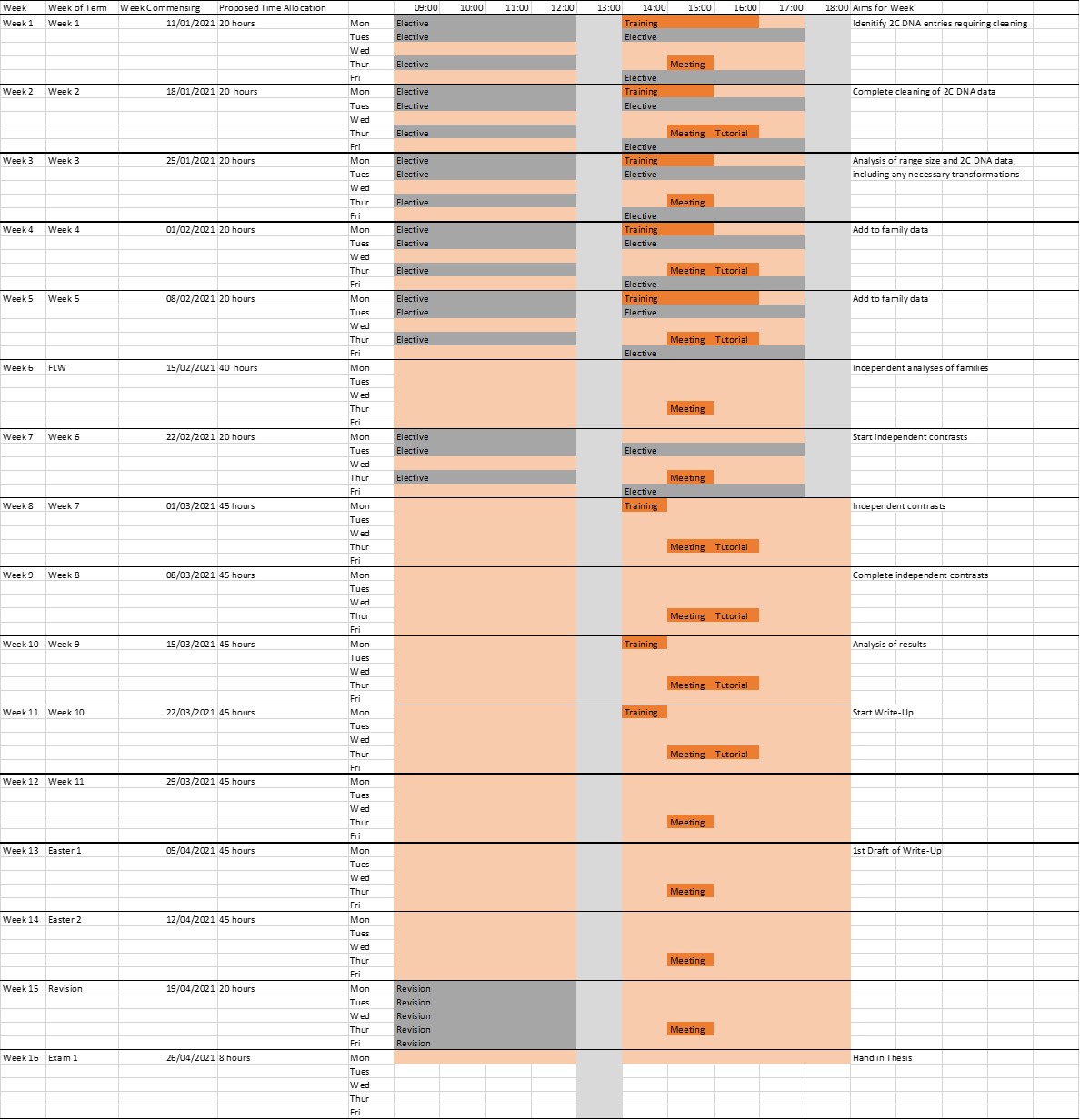


Figure 1- A flow chart exemplifying the sort of decision making that will be required during data analysis.

Figure 2- A table and Gant chart showing the proposed time allocation to the project. In the Gant chart: orange represents time allocated to the project (with dark orange representing time allocated to pre-organised training or meetings associated with the project), dark grey represents time allocated to other university study, light grey represents time not allocated to university study, and uncoloured boxes represent unallocated time.



**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. 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.

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

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.

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

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., 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.

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.

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.

Lee, S. and Kim, N. (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. Available at: <https://cvalues.science.kew.org/>. (Accessed: 10 January 2021).

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

Morris, J., *et al.* (2016). *Biology: How Life Works*. 2nd Edition ed: Macmillan Education.

POWO (2019). Plants of the World Online. Available at: <http://www.plantsoftheworldonline.org/>. (Accessed: 10 January 2021).

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.

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.

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.

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): 105-118.