**Taxonomic complexity in eyebrights (Euphrasia L., Orobanchaceae) and the British flora**

**Lay Summary:**

**Summary:**

Many plant groups are taxonomically complex with species that are difficult to distinguish. The main factors driving this complexity include apomixis, selfing, hybridisation, and polyploidy. Plant parasitism is a potential driving force of taxonomic complexity that has, however, been largely overlooked. In this thesis I use two main systems to explore these factors; the British flora and a hemiparasitic genus of plants, *Euphrasia*. The British flora is an excellent system with a wealth of large and comprehensive ecological and genetic data sets available, while *Euphrasia* is a tractable experimental system exhibiting rampant hybridisation, variation in ploidy level and mating system, and able to parasitise a wide range of plant species. The main aim of this thesis is to understand the role of hybridisation, polyploidy, and parasitism in driving taxonomic complexity in *Euphrasia* and the British flora. I first review the frequency and importance of cross ploidy hybridisation across plants, and based on a literature review and survey of the British flora find it to be more common than usually appreciated. Next, I investigate how hybridisation is affected by phylogenetic relationships and genetic distance between species across the British flora. I find that the probability of hybridisation is impacted mainly by parental genetic distance, ploidy level differences, and the extent of geographical overlap. Then, I investigate a single contact zone between a diploid and tetraploid species of *Euphrasia*, and find little evidence of contemporary hybridisation, however demographic modelling supports a model with low levels of gene flow. In the second part of the thesis, I use common garden experiments to understand the nature of species differences in the taxonomically complex, hemiparasitic genus *Euphrasia*. I show that firstly, traits used to identify species in *Euphrasia* are plastic, and change depending on the host plant species used. I go on to show that *Euphrasia* exhibit both conserved and host specific interactions across many different host species, which potentially reveals cryptic specialisation. This thesis shows how integrated analyses incorporating genetic and ecological data can be used to explore the many and diverse factors underlying taxonomic complexity in plants.

**Chapter 1 - General Introduction:**

**Taxonomic complexity in plants:**

Much of biology is built upon the idea that the diversity we see in nature can be categorised into discrete units, which at its heart is the concept of a species. Ever since this task has been embarked upon, difficulties have emerged as to the precise placement of certain species, mainly because the processes that give rise to species are in constant motion. Arguably among the most difficult groups to classify are collectively placed under the umbrella term of ‘taxonomically complex groups’ (TCGs). A TCG is here defined as one where it is difficult to categorise the biodiversity present due to underlying processes which blur species boundaries (ref). TCGs are problematic, not only in terms of classification but also as to how to best conserve them (refs). While TCGs are present across the whole tree of life (refs), TCGs are presented with remarkable frequency and abundance in plants, and particularly common in certain families such as Poaceae, Rosaceae, and Asteraceae (ref). Most TCGs are the result of plant populations that no longer undergo sexual, random mating (ref). The factors contributing to TCGs include selfing, apomixis (or agamospermy), hybridisation, and polyploidy. Selfing and apomixis disturb random mating and restrict gene flow between populations, and often interact with hybridisation and polyploidy, described below.

**The role of hybridisation in taxonomic complexity**:

Hybridisation, defined here as the exchange of genetic material between different species, is an important factor driving taxonomic complexity. The role of hybridisation is not simple however, as many different outcomes are possible. Hybridisation can be destructive, where rare species may lose their genetic integrity, resulting in populations of entirely hybrid genotypes (refs). On the other hand hybridisation can be creative, by allowing adaptive traits to move between species (ref), and can even lead to speciation (ref). Many TCGs involve hybridisation which blurs species boundaries, coupled with processes that may allow the hybrid derivatives to persist (ref). Examples include **agamic complexes**, in which hybridisation is combined with a mode of asexual reproduction (e.g. apomixis) to propagate lineages (ref), **polyploid complexes**, where the hybrid derivatives are sexual polyploids (ref), and **homogamic complexes** (or homoploid complexes, e.g. *Helianthus* (ref)), where hybrid derivatives are mainly diploid and isolated from parental progenitors ecologically (ref).

Some TCGs defy these three broad classifications, by combining properties of different species complexes. Good example of this are the genera *Sorbus* and *Euphrasia* in the British Isles. In *Sorbus*, apomictic species can iteratively generate novel lineages through hybridisation with sexual diploid progenitors, creating a complex reticulate network of related species which may themselves be both apomictic and sexual (refs). High levels of selfing, combined with differing ploidy levels and extensive hybridisation in *Euphrasia* has led to a vast array of taxon descriptions, many of which evade current taxonomic treatment (refs). Not only is hybridisation important in generating biological diversity and complexity, it is also a common phenomenon both geographically, and phylogenetically (refs). The clear and widespread abundance of this phenomenon means it has had, and continues to have, a profound effect on the evolution of plants – especially in conjunction with selfing, apomixis, and polyploidy.

**The role of ploidy variation in taxonomic complexity**:

Polyploidy is the condition where a cell contains more than two sets of chromosomes as a result of whole genome duplication (WGD), and is featured in almost all TCGs. The two major routes to polyploidy are either through WGD of a single species chromosome complement, known as autopolyploidy, or through hybridisation between two species followed by WGD, known as allopolyploidy (however, there are other mechanisms; see (Ramsey and Schemske, 1998)). Although the majority of plant species are diploid (~67%, (Rice et al., 2019), extensive variability in ploidy levels exist across all flowering plants, at all taxonomic levels (Kolar et al., 2017, Soltis et al., 2010). Both the spatial and phylogenetic distribution of ploidy variation are unlikely to be uniform due to climatic and clade specific effects on unreduced gamete formation, which is the main driver in the creation of polyploid organisms (Kreiner et al., 2017a, Bretagnolle and Thompson, 1995, Rice et al., 2019).

Polyploids drive taxonomic complexity by reducing gene flow between populations that differ in ploidy level, and polyploidy is also correlated with high levels of selfing and apomixis in some taxa. Recurrent polyploidisation can generate taxa of ever higher ploidy levels, some of which may persist to form new species, or backcross with parental species to form a complex reticulate phylogeny. Ploidy level also varies extensively even within species, and while allopolyploids tend to be classified as new species, autopolyploids generated at the population level are rarely warranted species status. This is despite autopolyploids generally showing strong geographical structuring, and can in some cases bearing differing morphologies (refs). Lastly, unusual cytogenetic features in some plant groups, for example unbalanced parental genomic contributions in permanent odd ploidy (pentaploid) roses (ref) and permanent translocation heterozygotes in evening primroses (ref), generate TCGs that have defied classification for over a hundred years.

**Novel features of taxonomic complexity; parasitism and plasticity**

Parasitic plants total around 1% of all angiosperm species, with some genera being particularly speciose such as *Pedicularis* (c.a. 650sp), *Euphrasia* (c.a. 246-350sp), and *Thesium* (c.a. 300sp) (refs). How different host species may influence parasitic plant phenotypes however, has been little explored in relation to taxonomic complexity (ref). Yet it is well known that different host species can dramatically impact the growth, development, and evolution of parasitic plant individuals (ref). There are two main ways in which the parasitic lifestyle may contribute to taxonomic complexity. Firstly, parasitic plant species present a plastic phenotype when utilising different host species, which may be substantial enough to confuse species identification. For example, a suite of traits used in species discrimination in the Orobanchaceae, including corolla length, node to first flower, and plant height all vary in relation to host quality. Secondly, differential host use can drive the evolution of cryptic taxa, which has been seen in the plant genus *Orobanche*, specialising on different host species (ref).

**The genus *Euphrasia***:

The genus *Euphrasia* (also known as eyebrights) are a large group of hemiparasitic plants in the Orobanchaceae numbering around 260 species worldwide (ref). The hemiparasitic condition is scattered throughout the plant phylogeny (ref) and is characterised by the ability of a plant to form a haustorium to draw nutrients and water from a host plant, while simultaneously being able to photosynthesise. *Euphrasia* were chosen as a focal group in the study of taxonomic complexity for a number of reasons. Firstly, the taxonomic complexity of *Euphrasia* in the UK is reflected in the hundreds of taxon names littered in the taxonomic literature (ref), and in the great diversity of morphologies found in the field. They also exhibit rampant hybridisation, with 69 hybrid combinations reported to date, and have high rates of selfing in small flowered species (ref). The possession of two different ploidy levels in the UK species (2n=2x=22 and 2n=4x=44) allows for the integration of ploidy into hybridisation and morphological data. In addition, being generalist parasites, *Euphrasia* can gain benefit from a wide variety of plant species which in turn impact the morphology, growth, and fitness of the parasitic *Euphrasia* plants. Lastly, from a horticultural point of view, *Euphrasia* are easily brought into cultivation, and their annual life history facilitates experimental work within the timeframe of a PhD.

**Questions, aims, and approaches:**

The main aim of my thesis is to investigate aspects of taxonomic complexity in postglacial *Euphrasia*, and across the British flora, with heavy emphasis on hybridisation, polyploidy, parasitism, and their interaction. The thesis is split into two parts which target different aspects of taxonomic complexity, and draw on broader themes in evolutionary biology. The first part concentrates on the contributions of hybridisation to taxonomic complexity in plants, with a case study in the genus *Euphrasia*. Here, I firstly ask how prevalent hybridisation is between plants that differ in ploidy level, and whether it is a significant evolutionary phenomenon (Chapter 2). The interaction between ploidy level and hybridisation is a poorly explored topic and scattered across the literature. I bring together a comprehensive list of examples, and synthesise current knowledge on this topic. After understanding this global variation, I concentrate on the British flora, and model the probability of hybridisation across the flora (Chapter 3). Significantly, this model includes phylogenetic relationships, and genetic distances based on the first complete DNA barcode dataset across a flora, which previous studies have not been able to comprehensively address (refs). Then, I focus on the promiscuous genus *Euphrasia* and ask whether we see evidence of hybridisation between divergent species of different ploidy at a fine spatial scale. Using reduced representation sequencing of genome wide markers (genotyping by sequencing; GBS), I use a combination of classical population genetic tools and demographic simulation, to understand the pattern of hybridisation in a *Euphrasia* contact zone (Chapter 4).

In the second part of the thesis, two novel features of taxonomic complexity - phenotypic plasticity and parasitism - are explored in the parasitic plant genus *Euphrasia*. I use common garden experiments to ask firstly how different host species affect the morphology of *Euphrasia* and the ability to discriminate between *Euphrasia* species (Chapter 5). This was investigated using a single species of *Euphrasia* grown across eight different host species, where various morphological traits of *Euphrasia* were quantified. Lastly, I look at the role of host parasite interactions to understand host specialisation in the genus, by growing multiple *Euphrasia* and multiple host species together (Chapter 6).

**Declaration:**

I hereby declare that the composition of, and the work contained within this thesis is my own work, unless explicitly stated otherwise.

[signature]

Max Brown

**Quotes:**

“If, again, references were given to the parasitism of Euphrasia, etc., how likely it would be that some young man would go on with the investigation; and so with endless other facts…”

Charles Darwin

“I remember my childhood names for grasses and secret flowers. I remember where a toad may live and what time the birds awaken in the summer -- and what trees and seasons smelled like - how people looked and walked and smelled even. The memory of odors is very rich.”

John Steinbeck

Cut text:

One widespread phenomenon (in 1% of known plant species; ref) is apomixis (or agamospermy), the asexual production of seed which can establish subtly different lineages from the sexual diploid progenitors. Only one lineage above genus level (*Houttuynia*;ref) is known to be persistently apomictic over a significant timescale, while all other apomictic taxa occupy the tips of the plant tree of life; examples include *Hieracium*, *Taraxacum*, *Rubus*, *Alchemilla*, and *Poa* (refs).