**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 (refs). Ever since this task has been embarked upon, difficulties have emerged as to how to delimit certain species, mainly because the processes that give rise to species are in constant motion (ref). Arguably among the most difficult groups to classify are collectively placed under the umbrella term of ‘taxonomically complex groups’ (TCGs; ref). 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). TCGs are present across the tree of life (refs), from fungi (ref), to fish (ref), and arthropods (ref). TCGs are remarkably frequent in plants however, and particularly common in certain families such as Poaceae, Rosaceae, and Asteraceae (ref). TCGs can be the result of plant populations that no longer undergo sexual, random mating (ref), however other TCGs are the result of phenotypic plasticity and recent speciation blurring species boundaries (refs). Factors commonly 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 (ref). Mating system shifts from outcrossing to selfing or apomixis, drive rapid reproductive isolation from progenitor lineages through founder effects, genetic drift, and selection of advantageous recessive alleles. Further, this leads to strong population structure, characterised many and varied subtle morphological changes between populations. Similar phenomena result from ploidy level variation (ref), however hybridisation can assist in moving genetic material between populations, which can produce yet more morphological variation (ref).

**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 (ref). Introgression can result from hybridisation, where there is a permanent incorporation of genetic material from one species in the genetic background of another (ref). 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, and can eventually lead to the extinction of the rarer species (refs). On the other hand hybridisation can be creative, by allowing adaptive traits to move between species (ref). In the extreme, new species can be formed in a process known as hybrid speciation (ref). Hybridisation which results in speciation has two main flavours: polyploid hybrid speciation where the hybrid species has duplicated its chromosome complement (allopolyploidy), and homoploid hybrid speciation where the parental species and the hybrid remain at the same ploidy level (refs). 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. A good example of this is the genus *Sorbus* in the British Isles (ref). In *Sorbus*, hybridisation, polyploidy, and apomixis have interacted to form numerous endemic species (ref). This process happens because there are few sexual diploid species (e.g. *S.torminalis, S.aria*, and *S.aucuparia*) which hybridise with apomictic polyploid derivatives of the species *S.aria* (ref). Apomictic *Sorbus* species require pollen to achieve successful asexual reproduction, however at low frequencies the pollen can fertilise the maternal embryo and this leads to a stable, new, polyploid, apomictic taxon (ref). In all, more than 31 apomictic species have arisen in this way (ref). The apomictic condition ensures that any new hybrid genotype is frozen in stasis, and can lead to complex reticulate evolutionary histories, which is seen in many other apomictic plant systems (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 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 involving plants. 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)). As polyploid individuals tend to be larger, more vigorous, and quicker growing, it is no surprise that many plant crop species are polyploid (e.g. wheat, rice, potato, maize refs). Although the majority of plant species are diploid (~67%, (Rice et al., 2019), extensive variability in ploidy levels exist across 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).

Polyploidy is a contributing factor to taxonomic complexity, with recurrent polyploid formation followed by hybridisation within and between ploidy levels being a main driver (ref). Further, there is evidence to suggest that polyploidy facilitates hybridisation as ploidy level increases, i.e. tetraploid x tetraploid crosses are more successful than diploid x diploid crosses (ref). Shifts in mating system from outcrossing to selfing or apomixis are also correlated with polyploid formation, due to the breakdown of self-incompatibility (ref), selection against of minority cytotype exclusion (ref), and the capability of polyploids to alleviate inbreeding depression (ref). Recurrent polyploidisation has been shown to 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 group (refs). Polyploid type (auto or allopolyploid) is important in determining their role in taxonomic complexity. Autopolyploids and allopolyploids are frequent in nature; current estimates for both kinds of polyploids are around 10% (refs). Autopolyploids form complexes within species where ecology can shape the distribution of cytotypes (ref), however these cytotypes are rarely considered species (ref). Allopolyploids contribute to taxonomic complexity by combining two distinct genotypes to form a usually distinct and isolated taxon from the parents (ref). Both kinds of polyploid are present within some TCGs (e.g. *Sorbus, Cochlearia*; refs).

Lastly, unusual cytogenetic features in some plant groups have generated TCGs that have defied classification for over a hundred years. Two examples are the dog roses in the section *Caninae* and the evening primroses in section *Oenothera* (refs). The dog roses present an unusual breeding system where the female parent contributes 3-5 copies of the genome and the male parent contributes only one (section *Caninae* species are usually pentaploid; refs). This results in the hybrids containing a genome which is mostly maternal, and has the added complication that reciprocal crosses entirely change the genomic constitution of the hybrid. Further, dog roses can hybridise frequently with roses from other sections of the genus, where chromosome numbers are likely to change in reciprocal crosses (refs). Another example of strange cytogenetic behaviour, but at the diploid level, are the evening primroses. Here, the species exhibit a breeding system where translocation hybrids and balanced lethals produce chains of chromosomes which are inherited as single units. While evening primroses breed true when selfing, when they hybridise different combinations of chromosome chains can produce completely new morphological taxa (refs).

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

Approximately 1% of all angiosperm species are parasitic, with some genera being particularly speciose such as *Pedicularis* (c.a. 650sp), *Euphrasia* (c.a. 246-350sp), and *Thesium* (c.a. 300sp) (refs). Parasitic plants are defined by the formation of a structure called the haustorium, which is used to extract water and soluble nutrients from the host plant (ref). The haustorium can attach either to roots (e.g. *Orobanche*; ref) or shoots (e.g. *Viscum* ref), or rarely the haustorium is present inside the plant itself (e.g. endoparasitic *Pilostyles*; ref). There are two types of parasitic plants – hemiparasites which retain photosynthetic competency, and holoparasites which are devoid of chlorophyll and entirely dependent on host plants (ref). Host species range and identity varies widely between parasitic plants, and some generalists can parasitise more than 100 host plant species (e.g. *Amyema miquelii*), while others specialise on a single or few host plants (e.g. *Epifagus virginica* on *Fagus grandiflora*; ref).

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 may show phenotypic plasticity when utilising different host species, which may be substantial enough to confuse species identification (ref). 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 (ref). Secondly, differential host use can drive the evolution of cryptic taxa, which has been seen in the plant genus *Orobanche*, which specialise on different host species (ref).

***Euphrasia* and the British flora**:

The genus *Euphrasia* (eyebrights) are a large group of 260 hemiparasitic plant species in the Orobanchaceae (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. 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.

[British flora section here]

**Questions, aims, and approaches:**

The main aim of my thesis is to investigate taxonomic complexity in *Euphrasia*, and across the British flora, with a particular focus on hybridisation, polyploidy, parasitism, and their interactions. 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).

**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).