Research Strategy

Hybrids often exhibit dysregulated development, and thus represent an unparalleled opportunity to explore how closely related species diverge in fundamental developmental processes. When these dysregulated hybrid phenotypes manifest in nature, they can also present barriers to reproduction, playing a key role in speciation[1]. Understanding how evolution shapes divergence in key reproductive and developmental processes also has implications for human health, as the evolutionary forces that shape divergence between species also shape variation within species, including variation in disease-associated alleles[2]. My lab leverages a model system for evolutionary genetics– the *Mimulus guttatus* species complex– to understand the genetic basis of reproductive barriers and the consequences of these barriers for the dynamics of gene flow. Much of my work involves dissecting the evolutionary drivers, and genetic basis of a fundamental reproductive barrier in both placental mammal and flowering plants; early onset hybrid inviability borne from inappropriate development of key nutritive tissues (e.g. placenta/endosperm).

Genomic imprinting and hybrid dysfunction

Intragenomic conflicts may play a central role in interspecific divergence, but empirical tests are needed[3]. A key source of conflict in viviparous organisms is conflict between maternal and paternal interests in resource allocation to offspring (i.e. parental conflict[4,5]). Parent conflict can arise in non-monogamous species because fathers are not equally related to all of the offspring produced, favoring the evolution of paternally derived alleles that solicit maternal resources. As maternal resources are limited and mothers are equally related to all of their offspring, natural selection will favor the compensatory evolution of maternally derived alleles that restrict resource allocation[4–6]. A genomic arms race between paternal excessive and maternal repressive alleles can ensue. Although theory dictates that maternal and paternal donors from the same population should be matched, crosses between populations that have differentially experienced this arms race can reveal mismatches in paternal excessive and maternal repressive alleles, resulting in parent-of-origin-specific growth defects. In flowering plants and mammals, hybrids often manifest parent-of-origin-specific defects in nutritive tissues that are essential for embryo development (i.e. endosperm or placental), causing embryo death. **Despite its commonality and potential role in diversification[7–9], the genes underlying these defects are unknown in any diploid system, and the role of parental conflict in driving their evolution is unresolved.**

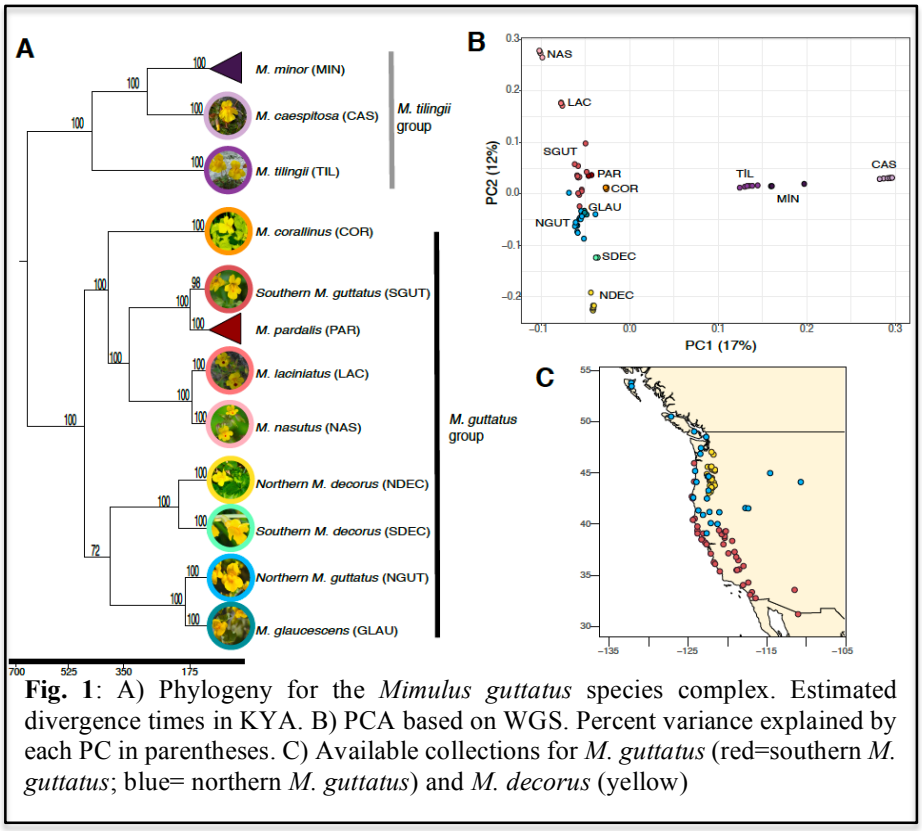

Fig. 1: A) Phylogeny for the *Mimulus guttatus* species complex. Estimated divergence times in KYA. B) PCA based on WGS. Percent variance explained by each PC in parentheses. C) Available collections for *M. guttatus* (red=southern *M. guttatus*; blue= northern *M. guttatus*) and *M. decorus* (yellow)

One mechanism for parent-of-origin-effect alleles is genomic imprinting, an epigenetic phenomenon whereby alleles exhibit parent-of-origin-specific expression[10–13]. Genomic imprinting is a cornerstone of placenta and endosperm development[10,11] that has evolved convergently in flowering plants and mammals[14,15]. Inappropriate endosperm/placenta development is associated with misexpression of imprinted genes[16–20], and imprinted genes often show signals of rapid evolution, consistent with an arms race[21,22]. In humans, misexpression of imprinted genes is associated with disease and infertility[23,24]. Thus, determining the evolutionary forces shaping imprinting has human health implications. However, producing thousands of offspring in controlled crosses in mammalian systems is challenging, which may limit our understanding of the role of parental conflict in imprinting divergence. Flowering plants represent a powerful alternative.

Mimulus is an ideal system to study whether parental conflict drives divergence in imprinting and hybrid developmental dysfunction (Fig. 1). The genus has substantial genetic and genomic resources and is ideal for genetic mapping. *Mimulus* are small, easy to grow, have short generation times (~2 months), and can produce thousands of

offspring per plant. Hundreds of highly inbred lines are available for crosses. A fully sequenced, assembled, and annotated genome exists for *M. guttatus*[25] and five other *Mimulus* species[26,27], and an ongoing collaboration with the Joint Genome Institute will yield 48 new reference genomes, including one for *M. decorus*; a focal species of my lab. *Mimulus* genomes are small (~400Mbp), making high coverage next-generation sequencing inexpensive. Consequently, hundreds of re-sequenced genomes are available for population-genomic analyses. In addition, over 1,000 highly polymorphic PCR-based markers[28], extensive RNAseq datasets, and many integrated genetic and physical maps are available. Stable transformation techniques (including CRISPR) are used regularly in *Mimulus* (see Yuan letter of support)[27,29–32]. Lastly, hybrid seed inviability is common, with over six independent incidences in the *M. guttatus* species complex and dozens more in the genus. Together, **the prevalence of hybrid seed inviability and exceptional genetic resources makes *Mimulus* an ideal system for characterizing the genetic and epigenetic basis of hybrid dysfunction.**

Recent Progress

Development of a model system: I previously described a new species– *M. decorus*– in the model group– *M. guttatus* species complex. Much of my work has sought to describe the evolutionary relationships, reproductive isolation, and adaptation in this species relative to the rest of the species complex[33–35]. *Mimulus decorus* is a montane perennial herb found at mid-to-high elevation throughout the cascade mountains in Oregon and Washington[35]. I have amassed >1,000 maternal families from 38 sites throughout its known range (Fig.1), including ~350 maternal families from 7 contact zones between *M. decorus* and *M. guttatus*. In addition, I have created over a dozen inbred lines, which has facilitated the generation of several finescale genetic maps. I have also re-sequenced 25 lines; the first re-sequencing data from this species[34]. Combined with the rich genetic and genomic resources for the syntenic and closely related *M. guttatus*, this work has created the necessary genetic tools to dissect the genetic basis of reproductive isolation in this group.

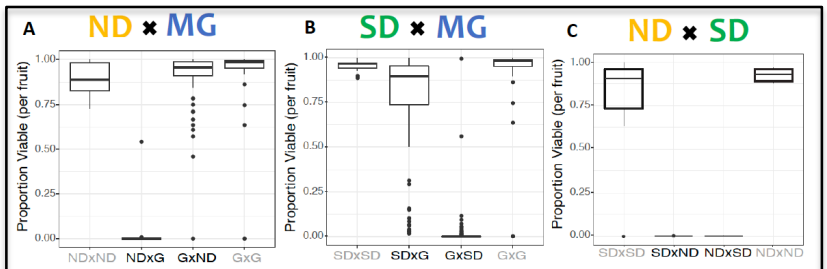

Fig. 2: (A,B,C) *M. guttatus* (MG) exhibits oppositely asymmetric hybrid seed inviability with northern and southern clades of *M. decorus* (ND and SD, respectively)

Hybrid seed inviability evolves rapidly: I performed reciprocal crosses between samples from 20 populations of *M. decorus* with 25 populations of *M. guttatus* to characterize the extent and diversity of reproductive isolation throughout the ranges of both species ([34], unpublished). Paired with range-wide re-sequencing, I discovered that *M. guttatus* and *M. decorus* are genetically distinct but recently diverged (~230KYA) and reproductively isolated via hybrid seed inviability[34]. At least two independent incidences of hybrid seed inviability have evolved in this group, as distinct, diploid clades of *M. decorus* (e.g. 'northern' and 'southern') show oppositely asymmetric hybrid seed inviability with *M. guttatus,* (and nearly complete hybrid seed inviability with each other[34]; Fig.2). Patterns of hybrid seed inviability conform to the predictions of parental conflict: (1) reciprocal F1s show parent-of-origin effects on seed size caused by abnormal endosperm development, (2) the extent of reciprocal F1 size asymmetries is correlated with reproductive isolation, and (3) inferred differences in the strength of conflict between populations predict the magnitude and directionality of hybrid seed inviability in subsequent crosses. **Hybrid seed inviability has rapidly and repeatedly evolved in this group, making it an unprecedented system to further investigate whether divergence in imprinting underlies hybrid incompatibility, and whether parental conflict is driving this divergence.**

Genetic Architecture of Hybrid Seed Inviability: Using a unique backcross design, I have begun to map the genetic basis of hybrid seed inviability between *M. guttatus* and northern *M. decorus*. I have detected a small number of parent-of-origin-specific loci that strongly influence the probability of seed survival (3 maternally- and 5 paternally- acting QTL; Fig.3). Each of these QTL reduces the probability of survival by 25-50% when maternally or paternally inherited, but not vice versa (Coughlan unpublished), which strongly suggests that genomic imprinting underlies this barrier. Several of these QTL contain promising candidate genes; including *SWN* (LG4) and *FIE* (LG13), both imprinted members of the FIS-PRC2[36,37] (the regulatory complex that creates genomic imprints in plants), as well as *EDE1* (LG13; an imprinted gene essential for *Arabidopsis* endosperm development[38]).

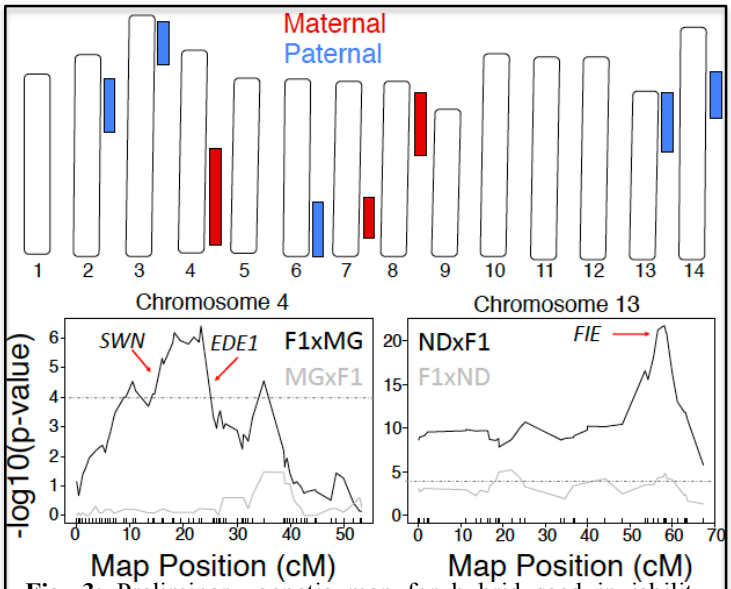

Fig. 3: Preliminary genetic map for hybrid seed inviability between northern *M. decorus* and *M. guttatus*. Top: linkage map with parent-of-origin QTLs and 95% CI. Bottom: Scan of a maternal and paternal QTL. Grey vs black lines depict patterns of segregation distortion for reciprocal crosses. Dotted lines depict significance thresholds for segregation distortion. Red arrows show locations of candidate genes. Vertical lines represent marker positions.

Hybrid zone dynamics: Introgression is ubiquitous across eukaryotes[39], and has significant evolutionary consequences, including contributing to both adaptive[40] and deleterious[41] genetic variation in humans. Determining which genomic regions are permissive to introgression can aid in our understanding of how effective reproductive barriers are in nature and identify sources of fitness-related genetic variation. I have begun to investigate the genomic landscape of introgression in *Mimulus*. I find that samples of *M. guttatus* and *M. decorus* that co-occur exhibit significant introgression, while samples from populations that do not co-occur do not exhibit such signals. This suggests that introgression is common and contemporary. Additionally, phenotypic surveys of the proportion of viable seeds produced in natural contact zones reveal that individuals from the contact zones often produce a high proportion of inviable, putatively hybrid seeds, suggesting that hybridization is rampant. I have previously shown that hybridization in the context of mix-pollen fruits can have significant potential growth and developmental implications for intraspecific seeds, which may influence their probability of survival in nature[42]. In future work, I aim to explore these consequences more fully.

Overview of Future Research

My lab integrates quantitative and population genomics, field work, and developmental biology to understand the evolutionary drivers and genetic basis of reproductive isolation and adaptation using the *Mimulus guttatus* species complex as a model system. The goal of the proposed work is to leverage the exceptional diversity of hybrid seed inviability in a group of closely related species to understand the extent to which parental conflict has driven divergence in genomic imprinting between recently diverged species and whether hybrid seed inviability has evolved as a byproduct of this divergence. Additionally, I will use seed collections from 7 natural contact zones to assess the dynamics of hybrid seed inviability in nature, quantify introgression, and begin to categorize other incompatibilities in this system.

What is the extent of divergence in genomic imprinting among closely related species?

Divergence in genomic imprinting may play a key role in plant and mammalian genome evolution and speciation[5,16–18,43–52], but little is known about the extent to which closely related species are divergent in imprinted expression and how natural selection has shaped this divergence. My group is working to (1) quantify the extent of divergence in genomic imprinting among *M. guttatus* and both northern and southern clades of *M. decorus*, (2) assess the role of parental conflict in driving divergence in imprinted expression, and (3) characterize the epigenetic landscape of divergent imprinted expression.

Catalogue imprinted expression among three closely related species: To explore imprinting divergence in *M. guttatus* and each clade of *M. decorus*, I will quantify patterns of imprinted expression within each species by performing reciprocal intraspecific crosses using distinct inbred lines from each species (Fig. 4). Using RNAseq, I can quantify expression in dissected endosperm across development. I will then identify genes that exhibit maternal or paternal bias in both reciprocal intraspecific crosses using linear mixed models[53]. Imprinted expression can be differentiated from allele-specific expression, as alleles exhibit biased expression regardless of parent of origin under the latter (Fig.4). A subset of putatively imprinted genes will be validated with qPCR to determine a false positive rate. At 50x coverage, power analyses indicate that I will be able to detect >95% of genes that show a parental bias of 20% or more from bi-parentally expressed expectations (a common threshold for imprinting[19,20,47,54–56]). By comparing imprinted expression

between species, I can quantify the extent of imprinting divergence across development. By repeating the above protocol in hybrid seeds, I can compare patterns of imprinted expression between hybrids and parents and identify genes that are differentially imprinted within species, but show abnormal expression in hybrids. These genes may comprise promising candidate genes for functional work.

Assess patterns of natural selection on imprinted genes: I will assess whether differences in gene expression are compatible with seletively neutral evolution, or instead suggests the action of natural selection, by comparing genetic and phenotypic divergence. Specifically, I will use a Qst-Fst approach[57] and phylogenetic comparative methods[58,59]. To assess whether parental conflict is driving this divergence, I will build on previous population genomic work that suggested that northern and southern clades of *M. decorus* and *M. guttatus* vary in their histories of parental conflict, wherein northern *M. decorus* has experienced the weakest history of conflict, and southern *M. decorus* has experienced the strongest[34]. **This natural continuum allows me to test whether species with a history of stronger conflict exhibit more imprinted genes and/or greater imprinted gene expression, as predicted by the parental conflict hypothesis**[47].

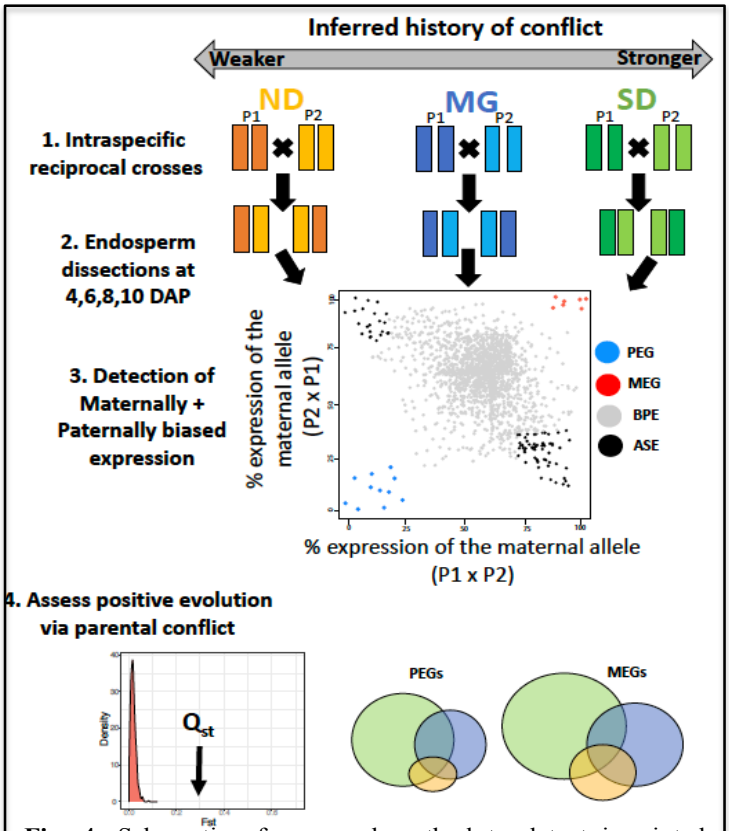

Fig. 4: Schematic of proposed method to detect imprinted expression in each clade of *M. decorus* and *M. guttatus*. (1) Use inbred lines of each species to create reciprocal intra-specific crosses. (2) Dissect and sequence endosperm at key developmental stages of seed development. (3) Detect genes that exhibit maternally or paternally biased expression in both crosses using linear mixed models (red; MEGs and blue; PEGs, respectively, versus ASE or bi-parentally expressed genes (BPE)). **Under a model of parental conflict, I predict that expression divergence will evolve more rapidly than neutral genomic divergence and that the number of MEGs and PEGs will be positively related to the inferred history of parental conflict.**

Characterize the epigenomic landscape of imprinted expression: In plants and animals, CpG, CHH, and CGH methylation plays an important role in modulating imprinting[10,12]. Although the epigenomic signatures of seed development have been explored, the ways in which epigenetic profiles diverge between species to create differential imprinting are poorly understood. To identify the epigenetic mechanisms underlying imprinting and discern how imprinting divergence occurs between closely related species, I will use bisulfite sequencing on endosperm to quantify endosperm-specific methylation profiles for *M. guttatus* and both clades of *M. decorus*. Methylation marks are strongly correlated with the presence of transposable elements (TEs)[60]. Using *de novo* genome assemblies of all three species, I will characterize TEs[61] and assess associations between TE proximity, imprinted expression, and methylation status to explore the epigenomic landscape of imprinting.

What is the genetic basis of hybrid seed inviability?

Despite the prominence of dysregulated endosperm/placenta in flowering plant and mammalian hybrids[9,16–18,43,45,46,48,50–52,62], few studies have mapped its genetic basis[16,17,19,51,63]. I will use genetic mapping to (1) test whether hybrid seed inviability is caused by nuclear parent-of-origin effect alleles, as would be expected if divergence in imprinting underlies this barrier, (2) assess the extent of overlap in alleles causing hybrid seed inviability between independence incidences as a measure of the degree of genomic parallelism in conflict-driven evolution, and (3) in tandem with the expression work, identify candidate genes for functional work. This work will be the first to identify the genes involved in early onset hybrid inviability borne from inappropriate endosperm/placenta development.

Assess the genetic basis and the extent of genetic parallelism among independent incidences of hybrid seed inviability. Currently, I am generating the genetic materials needed to map the genetic basis of hybrid seed inviability between *M.*

guttatus and each clade of *M. decorus* using segregation distortion mapping (Fig.5). To do this, I have created inbred lines for each species (<1% residual heterozygosity). I am creating reciprocal F1s between each pair of species, then backcrossing reciprocal F1s in both directions to both parents to create 8 mapping populations per species pair (Fig.5). As hybrid individuals that have inherited incompatible combinations of hybrid seed inviability alleles will perish, the surviving offspring will exhibit a deficit of the minor parent allele at regions of the genome involved in hybrid seed inviability, but no distortion in regions that are not involved in hybrid seed inviability. As F1s are used as both maternal and paternal donors, **this reciprocal F1 backcross uniquely allows me to identify alleles with parent-of-origin effects on segregation distortion**. Under a model of imprinting, I expect to find QTLs that show distortion when maternally or paternally inherited, but not vice versa (Fig.5). In addition, I will perform fine mapping by generating 2,000 F2 hybrids, identifying informatively recombinant individuals using PCR- based markers that flank each QTL interval, and sequencing those recombinant F2s using low-coverage (e.g. ~5x) whole-genome re-sequencing to provide a fine-scale map of recombination breakpoints in each informative F2. Low-coverage whole-genome sequencing can yield reliable ancestry assignment[64], which will not be difficult in this scenario, as these hybrids are born from a cross between divergent lines (D_{xy}=3.3-4.2% in all pairwise comparisons of *M. guttatus* and each clade of *M. decorus*). I will then backcross informative recombinants in each direction to each parent and quantify patterns of hybrid seed inviability to narrow down QTL intervals. Given the scale of recombination in *Mimulus*[25] and the scale with which recombination breakpoints can be mapped with low-coverage whole genome re-sequencing data (e.g.[65]), 2,000 F2 individuals should give excellent map resolution (~20 recombinant F2s per cM, with ~23 (max of 50) genes/cM). By comparing the extent of QTL overlap among independent incidences of hybrid seed inviability to a null expectation based on permutated maps we can quantify how repeatable conflict-driven evolution; a fundamental but unexplored question in evolutionary biology.

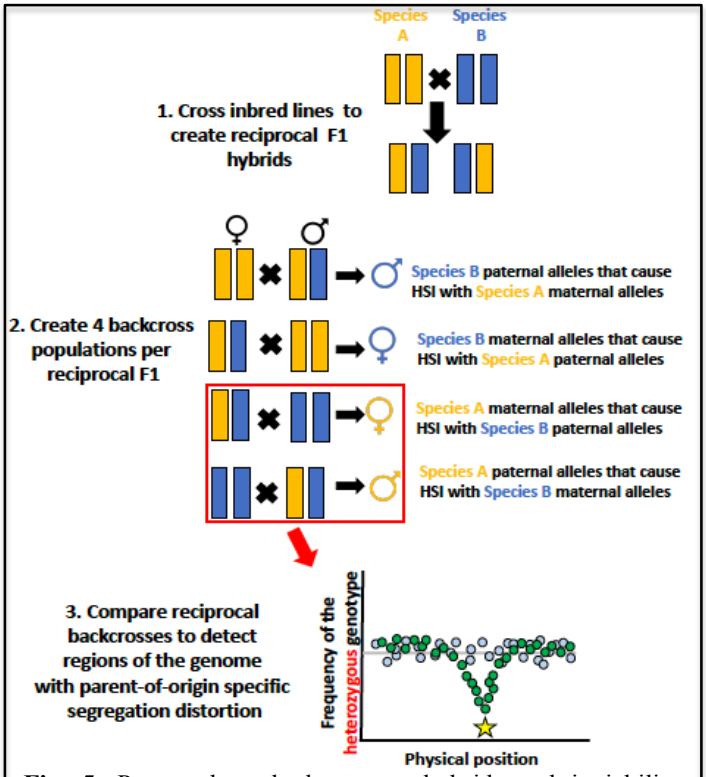

Fig. 5: Proposed method to map hybrid seed inviability between *M. guttatus* and each clade of *M. decorus*. (1) Cross inbred lines to create reciprocal F1s for each species pair. (2) Backcross each reciprocal F1 in both directions to both parental species. (3) Assess patterns of segregation distortion across the genome, with particular focus on parent-of-origin specific segregation distortion, indicative of imprinted loci. **Under a model of genomic imprinting, QTL will show parent-of-origin specific effects**

<u>*Functional validation of candidate genes*</u>: I will pinpoint candidate genes by identifying genes that are in the fine-mapped QTL interval and show differential imprinting among species and misexpression in hybrids. In *Maize* and *Arabidopsis* only 1.5-2.5% of genes show significant patterns of imprinting[53], and these tend to be evenly distributed across the genome (i.e. they are not genomically clustered, as in mammals[11]). For feasibility, I will prioritize QTLs with the largest effects on hybrid seed inviability. Under a model of imprinting, incompatibility results from inappropriate expression of normally imprinted alleles in hybrids. **If a loss of imprinting and subsequent overexpression of a gene causes hybrid seed inviability, then increasing expression of normally imprinted genes should induce hybrid seed inviability in a normally viable intraspecific cross, while reducing expression should rescue seed viability in a normally inviable interspecific cross.** To test this hypothesis, I will construct both overexpression lines with a seed-specific promoter and RNAi lines of the candidate genes to manipulate their expression in developing seeds (see Yuan letter of support). These techniques are regularly applied to *Mimulus*[27,29–31]. For all lines, I will verify successful transformation in *Agrobacteria* by sequencing, and again by qRT-PCR of endosperm in the final plants. I will then cross these lines in both directions to both parents, with the prediction that overexpression lines will induce inviability in previously compatible crosses, while RNAi lines will rescue viability in normally inviable crosses. Together, these

experiments can verify the genes causing hybrid seed inviability, and confirm that hybrid seed inviability is caused by their misexpression in hybrids; a central prediction of parental conflict.

Linking ancestry, incompatibility, and selection in the wild

Mimulus guttatus and both lineages of *M. decorus* commonly co-occur. I have begun to dissect the dynamics of introgression in this system, leveraging a small number of re-sequenced genomes. Re-sequencing of contact zones can not only help to explore introgression dynamics, and if/how hybrid seed inviability acts as a barrier to gene flow, but I can use patterns of ancestry disequilibrium (i.e. patterns of correlated ancestry) to identify new incompatibilities.

Identify the dynamics of introgression: I recently grew ~900 plants from 7 contact zones in a common garden. I self-fertilized each line and measured how many inviable seeds were formed as a proxy for how many incompatibility alleles are segregating within an individual. I found that the distribution of this phenotype was nearly continuous, suggesting that these sites contain individuals with variable ancestry proportions (Coughlan unpublished). I am currently sequencing a preliminary set of 100 individuals to assess the dynamics of gene flow, with the aim of sequencing the full panel in the future. Using ancestry HMMs I will identify the ancestry proportions for each contact zone as well as the distribution of haplotype lengths to assess the extent, timing and direction of introgression. I will also assess whether patterns of introgression correlate with particular genomic features, such as recombination rate, gene density, or divergence in parental populations, to quantify the extent of linked selection against minor parent ancestry[66–68]. Lastly, by assessing whether hybrid seed inviability QTL are less likely to introgress than other regions of the genome, I can begin to assess the efficacy of hybrid seed inviability as a barrier to gene flow.

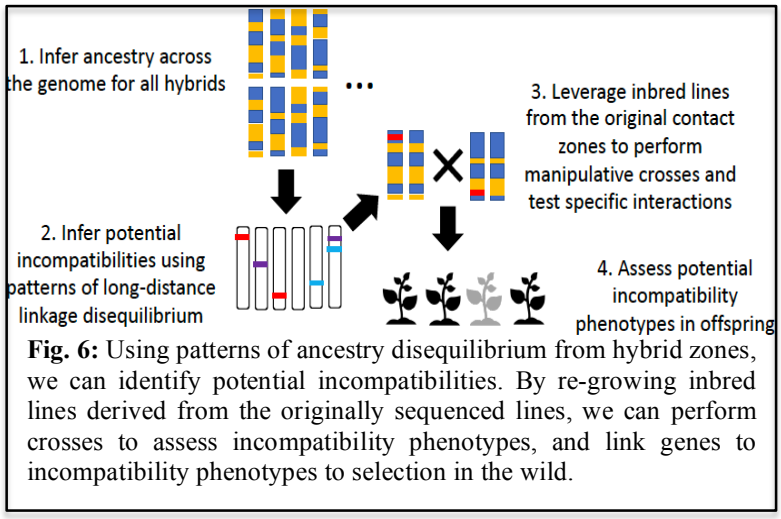

Fig. 6: Using patterns of ancestry disequilibrium from hybrid zones, we can identify potential incompatibilities. By re-growing inbred lines derived from the originally sequenced lines, we can perform crosses to assess incompatibility phenotypes, and link genes to incompatibility phenotypes to selection in the wild.

Identify genomic signals of incompatibility and leverage inbred lines to confirm incompatibility phenotypes: By leveraging patterns of ancestry and ancestry disequilibrium in natural hybrid zones, several research groups are making exciting progress on understanding the landscape of incompatibility[69–71]. One potential downfall to classical studies of ancestry disequilibrium is that these patterns can be caused by several processes, including traditional incompatibility, ecological selection, and assortative mating[72]. Pairing incompatibility phenotypes with specific loci that are found in ancestry disequilibrium is an important but often unachievable goal. A strength of our system is that all *Mimulus* are self-compatible, and thus the derivation of inbred, immortal lines is feasible. I have already begun making a set of 900 immortal lines from 7 contact zones between *M. guttatus* and each lineage of *M. decorus*. By using patterns of ancestry disequilibrium from the sequenced lines above, I can identify putative incompatibility alleles. Then, I can resurrect lines with known genotype at putative incompatibility loci, perform manipulative crosses, and identify potential incompatibility phenotypes. This powerful approach will allow me to connect genome-wide patterns of barriers to introgression with specific phenotypes, and likewise calculate selection coefficients for incompatibility loci in natural populations; a central but often unmet goal in speciation research.

Summary and Impact

The Coughlan lab leverages the diversity of the *M. guttatus* complex to characterize the role of parental conflict in generating reproductive isolation between closely related species and assesses the dynamics of introgression between strongly reproductively isolated species in nature. By integrating evolutionary and molecular genomics, I will characterize one of the most prevalent postzygotic reproductive barriers in flowering plants; from its molecular genetic basis to its role in species persistence. Additionally, I will quantify how repeatable evolution via conflict is; an important but understudied question. Understanding how evolution has shaped divergence in genomic imprinting, and the consequences of this divergence for reproduction, is a fundamental evolutionary question with important implications for human health and disease.