

Comparative Genomics of the Loss of Distyly and Self-Incompatibility in the Genus
Linum

By

Joseph Anton Maximilian Barham
Ecology & Evolutionary Biology, University of Colorado at Boulder

4 April, 2023

Thesis Advisor:

Nolan Kane, Ecology & Evolutionary Biology

Defense Committee:

Nolan Kane, Ecology & Evolutionary Biology
Pieter Johnson, Ecology & Evolutionary Biology
Dan Doak, Environmental Studies

Abstract

In all domains of life, inbreeding can be costly enough to favor the evolution of complex genetic mechanisms that prevent self-fertilization or inbreeding with close relatives. Conversely, the ability to self-fertilize can have advantages in some ecological contexts, such that this self-incompatibility has been gained and lost repeatedly across evolutionary history. An example of a recently-evolved self-compatible species from a self-incompatible species is *Linum lewisii*, a self-compatible, perennial, blue flax species in North America. Self-incompatibility is associated with a supergene cluster known as an S-locus in the genus *Linum*, and the breakup, change of function, or whole deletion of this supergene is assumed to be important in the history of *L. lewisii*. To investigate how this supergene has changed, I searched for the S-locus in the genome of *L. lewisii*. I have used sequence from a new genome assembly for *L. lewisii* and compared it with a previously published assembly for *L. tenue*, a distant self-incompatible congener with the hypothesized ancestral state of *lewisii*'s S-locus. I have used comparative genomics tools such as blast to identify the location of S-locus-associated genes in Lewis flax, and interpret their presence and absence in ecological and evolutionary contexts. Key findings include that a floral morphology-linked gene, TSS1, has been wholly deleted in Lewis flax and may explain some of its unique floral characteristics. My results also suggest that this S-locus supergene is likely ancestral to the genus *Linum*.

Background

Flax plants in the genus *Linum* have been used as an agricultural plant for fiber and oil since the dawn of agriculture (Fu *et al.*, 2012), and was used as a model system for studying complex sexual polymorphism in plants in Darwin's "On the existence of two forms, and on their reciprocal sexual relation, in several species of the genus *Linum*" in 1863. This classic biological research was focused on a polymorphism called distyly and the trait of self-incompatibility, *i.e.*: The complex set of traits that prevent self-fertilization in hermaphroditic species. *Linum* is subdivided into several sub-taxa called "sections." These include the section Adenolinum or Linum (generally the former to avoid confusion with the genus name), sometimes called the blue flaxes. Adenolinum includes common flax (*Linum usitatissimum*), the domesticated plant that produces linseed and linen. *Linum* is diverse and variable in many key traits, and Adenolinum includes many of these variations, such as sexual polymorphisms; variation in self-incompatibility; and variation in life cycle length, *i.e.*: perennial/annual (Reppinger, 2009). Thus, Adenolinum is a clade that is both useful as a model system, but also has the practical incentive that research in this clade can be directly related to an important agricultural crop (Fu and Allaby, 2009). The other section relevant to this study is known variably in the literature as Linoideae or Linopsis, and is one of the yellow-flowered clades. Linopsis does not feature any agronomically important species, but has been the subject of most of earlier work regarding the key varying traits in Adenolinum. My study will harness some of this earlier yellow flax research

and apply it to *Adenolinum*, in particular analyzing the genomic architecture of an unusual floral morphological trait of a North American *Adenolinum* species, *Linum lewisii* (common name blue prairie flax or Lewis flax) and compare it to a congener in *Linopsis*, *Linum tenue*, which exhibits the theorized ancestral state of these characteristics.

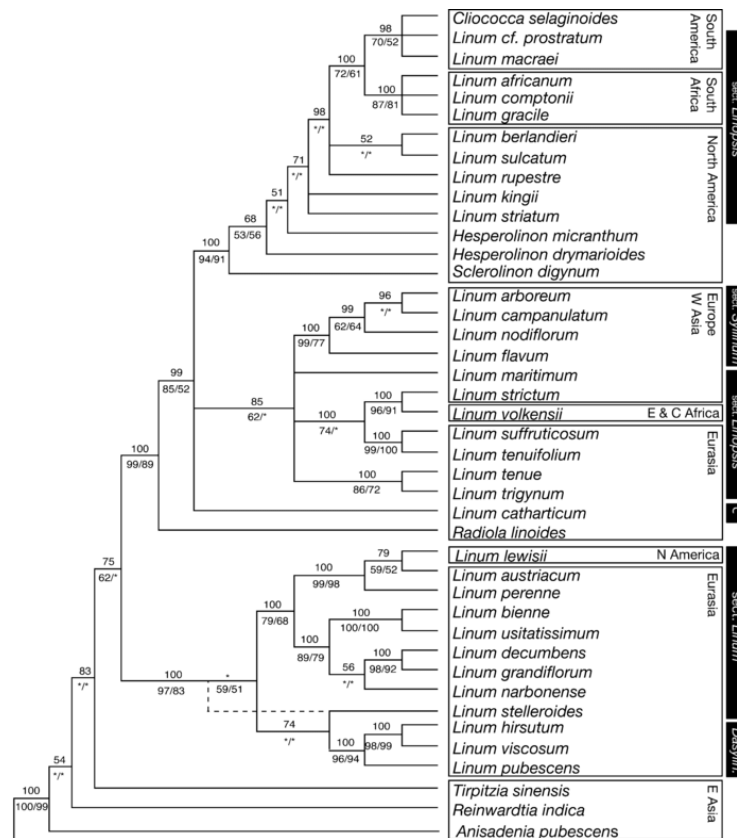


Fig.1: Excerpt of phylogeny of *rbcL* of *Linum* from McDill, 2009, showing position of *Adenolinum* (here referred to by its accepted alternate name “*Linum*”) and section *Linopsis*. Note that *Linopsis* is not monophyletic, whereas *Adenolinum* is generally accepted as being monophyletic.

The fundamental ecologically important difference between *L. lewisii* and its theoretical ancestral state is that Lewis flax can self-pollinate, while both *L. tenue* and the theoretical ancestral species have several key barriers to self-pollination. This is important because there are ecological and evolutionary advantages and disadvantages to self-pollination. Since self-pollinating involves the fertilization of an egg by a sperm from the same parent, it represents the most extreme form of inbreeding possible, and many offspring produced by self-pollination will be less fit than their out-crossed peers due to inbreeding depression (Charlesworth and Charlesworth, 1987; Charlesworth and Willis, 2009). Predictably, in many lineages across all domains of life, there are many strategies in hermaphroditic species to prevent resource wastage through self-fertilization.

Conversely, there can be advantages to self-fertilizing, especially in sessile organisms. In flowering plants, a plant with strong self-pollination barriers is at the mercy of chance to

reproduce successfully: If such a plant is not spatially proximate to a conspecific, its pollinator may not successfully transport pollen to it, or may do so in small or infrequent enough quantities that some of its eggs may remain unfertilized, thereby lowering fitness. In these edge cases, self-compatible plants have an advantage, as producing any number of inbred offspring is clearly favored over producing no offspring at all (Yuan *et al.* 2023). Thus, there are clear ecological contexts in which each mating strategy may be preferred: Obligate outcrossing will likely have an advantage in highly-competitive, relatively densely-populated areas by preventing resource wastage on relatively unfit offspring, whereas the capacity to self-pollinate may offer an advantage in more sparsely populated areas where mate availability is low.

In the context of section Adenolinum, most species are self-incompatible, and found in Eurasia (Replinger, 2009); however, *Linum lewisii*, the subject of this study, is a self-compatible North American species whose self-incompatible ancestor most likely crossed the Bering Sea land bridge during the last ice age (McDill *et al.*, 2009; Rogers, 1969). Mosquin posits in 1971 that this tendency toward self-compatibility in North American Adenolinum species is likely a consequence of adapting to the conditions of being an edge-species. However, the constant range-edge effect of sparse population density experienced by this ancestor as it colonized thousands of miles of land may have exerted a strong selective pressure on self-compatibility, leading to the present state. While this explains why self-compatibility may have been favored in the ancestor of blue prairie flax on a theoretical level, we must now examine the morphological and genetic traits that govern selfing. Self-incompatibility is not a single trait *per se*, but rather a complex mixture of traits that sum to a single effect.

Distyly, monostyly, and approach herkogamy are floral morphological characteristics that are a key part of self-compatibility status in *Linum*, either inhibiting or facilitating self-pollination (Barrett and Cruzan, 1994). Distyly is posited as the ancestral state of Adenolinum (McDill *et al.* 2009; Replinger, 2009), is the trait exhibited by *L. tenue*, and is the trait that interested Darwin and his contemporaries most, as it is a complicated sexual polymorphism. Lewis flax exhibits the somewhat simpler trait of approach herkogamy, and common flax exhibits monostyly. Each will be explained, along with its effects and values in varying ecological contexts.

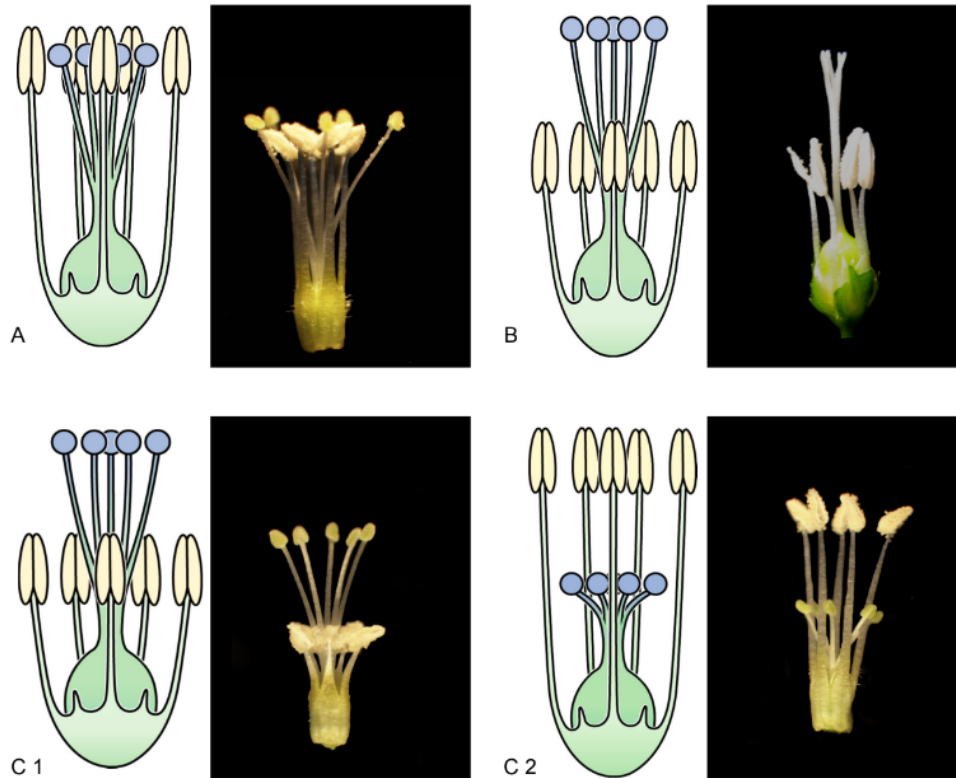


Fig.2: Monostyly (A), approach herkogamy (B), and distyly morphs thrum (C 2) and pin (C 1) which are found on different individuals within the same species. Note that pin and approach herkogamy are the same morphological configuration. From Reppinger (2009).

Monostyly represents the simplest case, in which male anthers and female stigma are held at the same level, and are thus spatially proximate. This trait is often associated with plants that commonly self-pollinate as a core component of their reproductive strategy (Ernst, 1954). Herkogamous plants are those in which the male and female reproductive parts are spatially separated from one another. This can reduce the chance of self-pollination without the need for evolving more complicated self-incompatibility barriers at a genetic level, though in self-incompatible plants it can still be valuable as it prevents pollen getting stuck on the stigma and being wasted (Webb and Lloyd, 1986). *L. lewisii* is a self-compatible plant exhibiting a form of herkogamy called approach herkogamy, in which the filaments are much shorter than the styles, causing the female stigma to be held above the male anthers.

Distyly is the most complicated and is section *Adenolinum*'s most common floral morphology, as well as the morphology exhibited by *L. tenue*. In distylous species, some individuals have long styles and short filaments, thus holding their female stigma above their male anthers (known as "pin" or "L-morph"), while other individuals in the same species have short styles and long filaments, and therefore hold their anthers above their stigma ("thrum" or "S-morph"). Pin-morph plants are thus morphologically identical to approach herkogamous plants, which can occasionally make identifications based on morphology difficult. In *Linum*,

distylous plants are usually both self-incompatible and intramorph incompatible, meaning that a thrum cannot typically successfully pollinate itself nor another thrum, and must pollinate a conspecific of the other morph (Barrett and Cruzan, 1994; Reppinger, 2009). Distyly's complexity and ecological importance is what has captivated researchers since Darwin; however, while these earlier biologists were impressively insightful in their predictions about distyly's genetic architecture and inheritance, it is only now, in the era of bioinformatics, that we can directly observe the genomes of these plants and test their hypotheses.

Supergenes are parts of the genome containing several individual genes that often code for complicated polygenic traits that need all of these genes to work properly. As such, recombination of these regions is often suppressed, and the region is passed down to offspring as a single unit, hence the name "supergene" (Thompson and Jiggins, 2014; Bateson and Gregory, 1905). After the invention of modern sequencing and computing, a supergene in the genus *Primula* was identified and characterized as being the locus of control for distyly and self-incompatibility in that genus (Huu *et al.*, 2016; Li *et al.*, 2016; Burrows *et al.*, 2017). In accordance with older literature predicting the existence of such genes, this supergene was named the "S-locus," as this was the name given to the theoretical loci that control for self-incompatibility. These studies have become the basis for further study of the genetic controls of distyly and self-incompatibility, such as recent studies in *Linum*.

Using similar methods as the *Primula* studies in the genus *Linum*, Gutiérrez-Valencia *et al.* found and characterized an S-locus supergene in the species *L. tenue* in 2022. In *L. tenue*, just as had earlier been found in *Primula*, individuals carrying the S-locus (genotype S/s) are thrum-type in morphology, while individuals without the S-locus (genotype s/s) are pin-type. Recall that because thrums are generally incompatible with other thrums in *Linum*, an S/S genotype is theoretically impossible, or at least highly unlikely. Architecturally, the S-locus is hemizygous in all characterized cases, meaning that it acts like an insertion on carrying S-type chromosomes, and is simply absent on non-carrying s-type chromosomes.

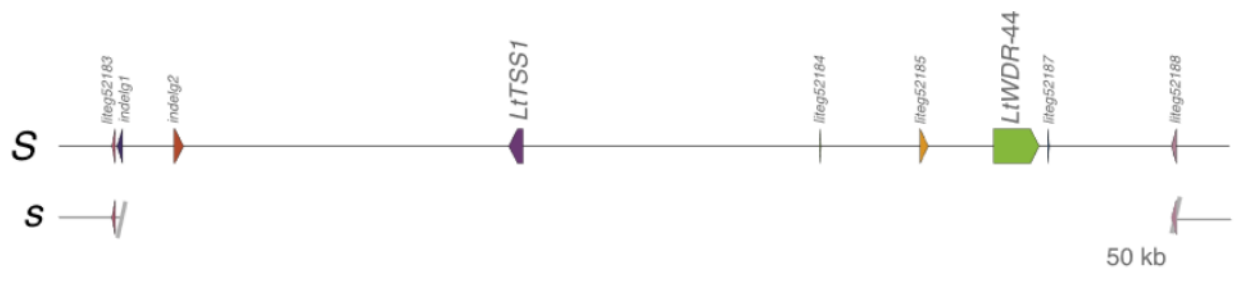


Fig.3: Map of the *L. tenue* S-locus. In S-genotype chromosomes of *L. tenue*, the S-locus is an insertion on chromosome 10. On s-genotype chromosomes, this supergene is "deleted", and the sequence simply continues on the other side. From Gutiérrez-Valencia (2022).

Introduction

Synthesizing all this information, we come to *L. lewisii*'s place in the story: *L. lewisii* is a self-compatible, approach herkogamous plant likely descended from a distylous, self-incompatible ancestor similar to *L. tenue* which lost its distyly and self-incompatibility during the crossing to the Americas, possibly due to the advantages of selfing imparted by edge-effects inherent to a range-expanding population. Because the S-locus is a complicated cluster of multiple genes working together, it seems far more likely that it has been lost independently several times than that it has arisen independently several times. Morphologically, *L. lewisii* is nearly indistinguishable from the pin-types of its close Eurasian relatives, but until the advent of bioinformatics, the mechanisms by which this transition occurred could not be directly studied. Now, with comparative genomics tools, we can test the various hypotheses of how Lewis flax may have lost its S-locus-associated ancestral traits:

- 1) Modern Lewis flax, with its pin-like morphology, is descended exclusively from a pin-type founder population that gained self-compatibility and intramorph compatibility and was thus favored as its range expanded.
- 2) Modern Lewis flax is descended from a more typical distylous founder population containing both pin and thrum individuals that evolved universal pin-like morphology through mutations in the S-locus.

In the first scenario, *L. lewisii* would be thought to have come to North America in a rare long-distance dispersal event from a *L. perenne*-like ancestor, leading to a small population of founders. If these individuals were coincidentally all pin-type, the only way they could reproduce would be through a gain of self-compatibility by some mutation. We would therefore expect the genome of *L. lewisii* to have no evidence of the S-locus in this scenario, as if modern blue prairie flax is descended only from pin-types, its ancestor should have had the “deletion” instead of the S-locus. This scenario would also easily explain the morphology of modern Lewis flax, as it is morphologically nearly indistinguishable from its pin-type distylous relatives. Thus, if no evidence of the S-locus were found in *L. lewisii*, it would be consistent with this first hypothesis. This first hypothesis is consistent with the pin-like morphology of all known Lewis flax populations, but has not been investigated genomically.

Alternatively, in the second scenario, we would expect that the S-locus be present, either in its entirety as a defunctionalized or refunctionalized supergene, or in fragments or clusters throughout the genome of *L. lewisii*. Some level of defunctionalization or refunctionalization would be expected of the S-locus in this scenario to allow self-compatibility, and to allow for selection on S-locus associated traits such as thrum-morphology (Baker, 1966). This could potentially be by a number of mechanisms, such as large-scale deletions from x-rays, unequal crossing-over, or through interactions with transposons such as the 7 already found in the S-locus of *L. tenue* by Gutiérrez-Valencia *et al.*. Inactivation of certain S-locus genes could induce a

pin-like morphology in thrum-genotype individuals (Ushijima *et al.* 2011). The presence of any S-locus genes in the genome of *L. lewisii*, intact or otherwise, would be evidence of this second hypothesis as it would imply that thrums were present in the founder population despite the current pin-like morphology.

Methods

I used a combination of Basic Local Alignment Search Tool (hereafter BLAST) analyses (Altschul *et al.* 1990) to find S-locus genes in Lewis flax. BLASTs function by receiving a “query” sequence you want to find, and a “subject” sequence in which to look for the query. There are several types of BLAST, in which the query and subject may be a nucleotide sequence file or a translated protein sequence file. Fundamentally, the goal of any BLAST is to try to identify where the subject and query “align,” and allow scientists to interpret the presence, absence, and structure of such alignments. As such, BLASTs are extremely useful in evolutionary genetics. In this study, I used two types of BLAST:

tBLASTn was my primary analysis tool. It translates the subject’s DNA sequence in all six reading frames and generates all possible proteins that DNA could code for. Then, it attempts to align these proteins to the query list of known protein sequences. In practice, this allows us to compare the known primary structure of proteins in the query to all possible primary structures of proteins in the subject and identify similar proteins between the two, with the important caveat that the subject’s possible proteins are not necessarily expressed, it merely demonstrates that the DNA sequence for those proteins exists. BLASTn was my secondary analysis tool, used to double-check surprising or ambiguous results. BLASTn then attempts to align DNA sequence from the query to DNA sequence in the subject, and thus simply shows DNA similarities and differences between the two. In all BLASTs, a location in which a similarity is found is known as a “hit.”

I performed several BLASTs: First, a tBLASTn in which the query was a list of the proteins in the S-locus of *L. tenue* (Gutiérrez-Valencia *et al.* 2022), and the subject was the entire draft genome of *L. lewisii* (Innes *et al.*, in prep), with output format 6. I used a quality filter of at least 40% sequence similarity and a bit score of at least 100 as the significance threshold.

Next, a BLASTn was performed with the *L. tenue* S-locus DNA sequence as the query and the entire Lewis flax genome as the subject. This was used to verify that the specific gene TSS1 was not present. I performed a subsequent BLASTn using the untranslated DNA sequence of the gene LgTSS1 from Ushijima *et al.*, 2011, as the query and the entire Lewis flax genome as the subject. Finally, a reciprocal BLAST similar to that done by Gutiérrez-Valencia *et al.* in 2022 was performed on the paralog of WDR-44 found in *L. tenue* to determine which Lewis flax copy correlated to which *L. tenue* copy. This was done by performing a second tBLASTn using the paralogs of WDR-44 from *L. tenue*’s genome as the query and the *L. lewisii* genome as the subject, thus giving data about which *L. lewisii* copy is most similar to the paralog WDR-44.

Results

Gene	Status in <i>L. lewisii</i>	Chromosome in <i>L. lewisii</i>	Position of best hit	% Identity of best hit	Bit Score of best hit
TSS1	Absent	N/A	N/A	N/A	N/A
WDR-44 S-locus gene	Present	5	31599032	57.868	211
WDR-44 Paralog*	Present	3	41932576	82.447	515
liteg52183	Present	3	69681807	40.741	108
indelg3	Multiple Hits	Several	N/A	N/A	N/A
indelg6	Insignificant	N/A	N/A	N/A	N/A
indelg7	Multiple Hits	Several	N/A	N/A	N/A
indelg8	Insignificant	N/A	N/A	N/A	N/A
indelg9	Insignificant	N/A	N/A	N/A	N/A
liteg52188	Present	3	69665885	91.558	289

*Data for this row taken from the reciprocal BLAST, while all other rows are from the original BLAST

The above table represents the results of the tBLASTn of all 9 S-locus associated genes that were identified by Gutiérrez-Valencia *et al.* in *L. tenue*, as well as the reciprocal tBLASTn of WDR-44's paralog, indicating their presence, absence, or the level of uncertainty associated with them in my analyses. Strikingly, TSS1 has no hits of any quality. Its absence was confirmed by *post hoc* BLASTn analyses, in which the query DNA sequence for TSS1 from both *L. tenue* and *L. grandiflorum* (Ushijima *et al.*, 2011) also failed to identify any trace of TSS1 in *L. lewisii*. Two genes, liteg52188 and liteg52183 are relatively proximate to one another on the 3rd chromosome of *L. lewisii*, though whether these genes are from the S-locus or from an ancestral paralog of the S-locus genes is difficult to determine with these data. Several genes are listed as “Multiple Hits” because many high-quality hits of them were found on several chromosomes, with no clear true location. Other genes are listed as being “Insignificant.” In these cases, many low-quality hits were found (as opposed to TSS1's no hits at all), but no hits passed the significance threshold of a bit score of 100.

The reciprocal tBLASTn performed on WDR-44 and its paralogs revealed that the S-locus copy of WDR-44 better correlates to Lewis flax's 5th chromosome, while the paralog copy better correlates to Lewis flax's 3rd chromosome. This is accomplished by comparing the quality of hits, and assuming that the highest quality hits correlate to genes with a common origin, as they should be the most similar by the basic principles of evolution.

Discussion

Broader evolutionary implications:

Having now shown evidence of the presence of S-locus genes in *L. lewisii*, my research supports the second and third hypotheses posited in the introduction, *i.e.*: That modern blue prairie flax has thrum ancestry. This rejects the hypothesis that a pin-type ancestor gained self-compatibility and gave rise to the morphologically-similar modern blue prairie flax lineage, and favors a somewhat more complicated history in which thrums' genetic data continued to exist, but the thrum's morphological characteristics were lost. There are many mechanisms that could explain this process, from partial recombination of the S-locus to defunctionalization of key genes in the S-locus, perhaps via transposon insertion (Vieira *et al.*, 2003; Vieira *et al.*, 2008; Berdan *et al.*, 2022), but precisely what happened will take more analysis to discover. In whatever case, these early self-compatible plants likely would have been disfavored in any context except those in which self-compatibility was strongly favored, as their floral morphology would likely have been affected given that many S-locus genes are directly involved with floral morphology (Gutiérrez-Valencia *et al.*, 2022; Ushijima *et al.* 2011). However, in the right contexts, destabilized self-incompatibility and distyly systems could introduce enough variation to allow for selection on and against these traits, leading to the monomorphic, self-compatible sexual system employed by modern North American flax (Baker, 1966). I intend to study the mechanisms of this transition in the near future by searching for evidence of recombination in S-locus genes, and by investigating the role of transposons in this story.

A final evolutionary consequence of these findings, caused by the relatively large evolutionary distance between *L. tenue* and *L. lewisii*, is that at least some of the S-locus architecture appears to be ancestral to *Linum*. These two plants' most recent common ancestor is generally accepted to be the most recent common ancestor of the entire genus *Linum* (McDill *et al.*, 2009). This specific S-locus architecture is complex enough that it seems highly unlikely that it would independently arise twice, and therefore it is likely to be ancestral to *Linum* in general. Given that there is evidence of certain genes and architecture stretching back even further (Gutiérrez-Valencia *et al.*, 2022; Tuskan *et al.*, 2006; Bredeson *et al.* 2016), it is possible that the S-locus of *Linum* may include features shared with other clades in the family Linaceae.

TSS1:

The absence of TSS1 is interesting in light of other S-locus genes being present in Lewis flax. In *L. grandiflorum*, the gene is associated with style-length and is only expressed in thrum

individuals (Ushijima *et al.*, 2011). The gene is also found in plants as distantly related as black cottonwood from the family Salicaceae (Gutiérrez-Valencia *et al.*, 2022; Tuskan *et al.*, 2006) and cassava from the family Euphorbiaceae (Gutiérrez-Valencia *et al.*, 2022; Bredeson *et al.*, 2016), indicating that this gene may be relatively ancient in the order Malpighiales as *Linum*'s family Linaceae is not particularly closely related to Salicaceae nor Euphorbiaceae (Endress *et al.*, 2013; Xi *et al.*, 2012). Thus, because it is specifically correlated with thrum-type style length, I hypothesize that the deletion or at least defunctionalization of TSS1 may be an important step in the transition from distyly/self-incompatibility to approach herkogamy/self-compatibility in *Linum*, and potentially in Malpighiales in general. Conversely, the recruitment of TSS1 to the S-locus may have also been an important step in the formation of S-loci of this order. The next step would be to search for TSS1 in other *Linum* or Malpighiales where we see examples of distylous, self-incompatible plants and close relatives that have secondarily lost these traits to see whether this loss or defunctionalization of TSS1 is a pattern among secondarily self-compatible and monomorphic plants in these clades.

WDR-44:

WDR-44 has two paralogs in both *L. lewisii* and *L. tenue*. Gutiérrez-Valencia *et al.* demonstrated that the copy found outside the S-locus is likely an ancestral copy that predates the S-locus. The reciprocal BLAST revealed that the copy on chromosome 3 is most likely the paralog copy, while that on 5 is likely the S-locus copy. It is interesting that the ancestral copy is found on the chromosome with the other S-locus genes, but at this time I do not have an idea as to why this has occurred beyond chance. According to Ushijima *et al.* and Gutiérrez-Valencia *et al.*, WDR-44 is mostly thrum-expressed, and involved in pollen development. Whether WDR-44 is still expressed in *L. lewisii* is not yet known, and this will be an important step in determining its importance and place in Lewis flax's evolutionary history; however, given that it is important in pollen development, and pollen development is important in self-incompatibility (Silva and Goring, 2001), WDR-44 could be an key gene in the evolutionary history of distyly in *Linum*.

liteg52183 and liteg52188:

The exact function and expression of these genes is currently unknown (Gutiérrez-Valencia *et al.*, 2022). Having identified the locations of where these genes are in blue prairie flax, the next logical question is what each gene's function might be and where and when in the plant's body it is expressed. Currently, of the 9 S-locus associated genes found in the Gutiérrez-Valencia *et al.* study, only WDR-44 and TSS1 have information about their expression and function known. A transcriptomic analysis on the expression and differential expression of all these genes in *L. tenue* and/or *L. lewisii* would shed important light onto how these genes may be important to the history of distyly. It is also notable that these genes flank the S-locus in *L. tenue*, and that *liteg52183* and a truncated version of *liteg52188* are present in the pin genome. These genes are proximate to one another on *L. lewisii*'s chromosome 3, which may indicate some retention of the structure of the S-locus in that area. Low-quality BLAST hits for other

S-locus features are present in that same region as well, lending further impetus to look more closely: A nucleotide alignment of this region to the S-locus of *L. tenue* could be valuable in analyzing whether this cluster truly is a fragment of the S-locus.

indelg3 and indelg7:

These genes are listed as “Multiple Hits” because many high-quality hits of them were found on several chromosomes, and I was, as of the time of writing, unable to determine whether there was a clear true location with these analyses. It may be possible to develop a better methodology of screening, and I will attempt to do so in the future to see whether I can find conclusive evidence on these genes’ locations, or rigorously demonstrate their absence. What can be said from these data is that these genes appear to be highly repetitive in nature, which may be important to their function.

indelg6, indelg8, and indelg9:

These genes are listed as “insignificant” because none of their hits passed the significance threshold of a bit score of 100. This threshold is arbitrary, and some did have bit scores that approached this threshold. Therefore, it is difficult to say whether the gene might be present or absent based on these data, as these hits may be coincidence (sequence that appears similar, but are not the same gene), or they may be real but difficult to detect. Either way, in these cases the gene as seen in *L. tenue* could not be found in *L. lewisii* based on the thresholds set up *a priori*. Interestingly, some hits were proximate to the liteg52183 and liteg52188 on the 3rd chromosome, and may be worth a second pass as this may point to that cluster being a relatively complete, if degraded, copy of the S-locus. As mentioned above, a nucleotide alignment of this cluster would be useful.

Transposons:

It is worth mentioning that 7 transposons, genes that stochastically “jump” around the genome, were also listed in the Gutiérrez-Valencia *L. tenue* S-locus, and that these were originally in the tBLASTn analysis; however, they were omitted from this paper because each would be considered “inconclusive” or “insignificant” from the level of analysis already performed. More sophisticated methodology may be able to glean information from them in the future, which will be important as transposons likely have an important part to play in the history of S-loci, as insertions of transposons at key locations could be a mechanism by which the S-locus was defunctionalized in the first place. Thus, while this study does not have the depth necessary to speculate on transposons at this time, they are a key missing piece of the puzzle that I intend to fill in in the future.

Future Work:

Beyond the work outlined above, a key future step to understanding the breakdown of distyly and self-incompatibility in Lewis flax (and *Linum* in general) would be assembling and

annotating a distylous *Adenolinum* plant or, alternatively, assembling and annotating a self-compatible and monomorphic close relative of *L. tenue*. *L. tenue* is an imperfect species for comparative genomic analysis of this nature because it is relatively distantly related to Lewis flax, despite being a congener (Replinger, 2009; McDill *et al*, 2009). An advantage of this distance is that we now have evidence that the S-locus may be ancestral to genus *Linum*; however, using a close relative of Lewis flax's genome, such as *L. perenne*, would give better data with which to compare the S-locus' genes and the mechanisms of its decay. Since *L. lewisii* and *L. perenne* diverged so recently in evolutionary history, it is probable that genetic differences beyond those associated with self-incompatibility and distyly are minimal. This comparison would give a remarkable view into the breakdown of distyly and self-incompatibility, far beyond what is possible using existing genomic data of such distant relatives. In order for this to occur, however, it will first be necessary to assemble and annotate the genome of one such species, a project that would be both a large amount of work and extremely informative on many other subjects beyond the S-locus.

Bibliography

- Altschul, S.F., W. Gish, W. Miller, E.W. Myers, D.J. Lipman (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*. **215**: 403-410
- Baker, H.G. (1966). The Evolution, Functioning, and Breakdown of Heteromorphic Incompatibility Systems. I. The Plumbinaceae. *Evolution*. **20**: 349-368
- Barrett, S.C.H., and M.B. Cruzan (1994). Incompatibility in Heterostylous Plants. In Williams, E.G., A.E. Clarke, R.B. Knox (Eds.), *Genetic Control of Self-Incompatibility and Reproductive Development in Flowering Plants: Advances in Cellular and Molecular Biology of Plants*. **2**. Springer, Dordrecht.
- Bateson, W., and R.P. Gregory (1905). On the Inheritance of Heterostylism in *Primula*. *The Royal Society of London*. **76**: 581-586
- Berden, E.L., T. Flatt, G.M. Kozak, K.E. Lotterhos, B. Wielstra (2022). Genomic Architecture of Supergenes: Connecting Form and Function. *Philosophical Transactions of the Royal Society B*. **377**: 20210192

- Bredeson, J.V., J.B. Lyons, S.E. Prochink, G.A. Wu, C.M. Ha, E.E. Gonzales, J. Grimwood, J. Schmutz, I.Y. Rabbi, C. Egesi, P. Nauluvula, V. Lebot, J. Ndunguru, G. Mkamilo, R.S. Bart, T.L. Setter, R.M. Gleadow, P. Kulakow, M.E. Ferguson, S. Rounsley, D.S. Rokhsar (2016). Sequencing Wild and Cultivated Cassava and Related Species Reveals Extensive Interspecific Hybridization and Genetic Diversity. *Nature Biotechnology*. **34** 562-570
- Burrows, B., and A. McCubbin (2018). Examination of S-Locus Regulated Differential Expression in *Primula vulgaris* Floral Development. *Plants*. **7**: 38
- Charlesworth, B., and J.H. Willis (2009). The Genetics of Inbreeding Depression. *Nature Review Genetics*. **10**: 783-796
- Charlesworth, D., and B. Charlesworth (1987). Inbreeding Depression and Its Evolutionary Consequences. *Annual Review of Ecology and Systematics*. **18**: 237-68
- Darwin, C. (1863). On the existence of two forms, and on their reciprocal sexual relation, in several species of the genus *Linum*. *Botanical Journal of the Linnean Society*. **7**: 69-83.
- Endress, P.K., C.C. Davis, and M.L. Matthews (2013). Advances in the Floral Structure Characterization of the Major Subclades of Malpighiales, One of the Largest Orders of Flowering Plants. *Annals of Botany*. **111**: 969-985
- Ernst, A. (1954). Self-Fertility in Monomorphic *Primulas*. *Genetica*. **27**: 391-448
- Fu, Y.B., A. Diederichsen, and R.G. Allaby (2012). Locus-Specific View of Flax Domestication History. *Ecology and Evolution*. **2**: 139-152
- Fu, Y.B., and R.G. Allaby (2009). Phylogenetic Network of *Linum* Species as Revealed by Non-coding Chloroplast DNA Sequences. *Genetic Resources and Crop Evolution*. **57**: 667-677
- Gutiérrez-Valencia, J., M. Fracassetti, E.L. Berdan, I. Bunikis, L. Soler, J. Dainat, V.E.

- Kutschera, A. Losvik, A. Désamoré, P.W. Hughes, A. Foroozani, B. Laenen, E. Pesquet, M. Abdelaziz, O.V. Pettersson, B. Nystedt, A.C. Brennan, J. Arroyo, and T. Slotte (2022). Genomic Analyses of the *Linum* Distyly Supergene Reveal Convergent Evolution at the Molecular Level. *Current Biology*. **32**: 4360-4371
- Huu, C.N., C. Kappel, B. Keller, A. Sicard, Y. Takebayashi, H. Breuninger, M.D. Nowak, I. Bäumle, A. Himmelbach, M. Burkart, T. Ebbing-Lohaus, H. Sakakibara, L. Altschmied, E. Conti, and M. Lenhard (2016). Presence Versus Absence of *CYP734A50* Underlies the Style-length Dimorphism in Primroses. *eLife*.
- Innes, P., B. Smart, C. Kane, *et al.* (in prep) Chromosome-Scale Genome Assembly of Lewis Flax (*Linum lewisii* Pursch.)
- Li, J., J.M. Cocker, J. Wright, M.A. Webster, M. McMullan, S. Dyer, D. Swarbreck, M. Caccamo, C. van Oosterhout, P.M. Gilmartin (2016). Genetic Architecture and Evolution of the S Locus Supergene in *Primula vulgaris*. *Nature Plants*. **2**: 16188
- McDill, J., M. Replinger, B.B. Simpson, J.W. Kadereit (2009). The Phylogeny of *Linum* and Linaceae Subfamily Linoideae, with Implications for Their Systematics, Biogeography, and Evolution of Heterostyly. *Systematic Biology*. **34**: 386-405
- Mosquin, T. (1971). Biosystematic studies in the North American species of *Linum*, section Adenolinum (Linaceae). *Canadian Journal of Botany*.
- Replinger, M. (2009). Phylogenie und Biogeographie der Linoideae (Linaceae) - Evolution von Heterostylie/Homostylie in *Linum*. *Johannes-Gutenberg-Universität in Mainz*.
- Rogers, C.M. (1969). Relationships of the North American species of *Linum* (flax). *Torrey Botanical Society*. **96**: 176-90.
- Silva, N.F., D.R. Goring (2001). Mechanisms of Self-Incompatibility in Flowering Plants.

Cellular and Molecular Life Sciences. **58**: 1988-2007

Thompson, M.J., and C.D. Jiggins (2014). Supergenes and Their Role in Evolution. *Heredity*.

113: 1-8

Tuskan, G.A., S. Difazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, S.

Ralph, S. Rombauts, A. Salamov, J. Schein, L. Sterck, A. Aerts, R.R. Bhalerao, R.P.

Bhalerao, D. Blaudez, W. Boerjan, A. Brun, A. Brunner, V. Busov, M. Campbell, J.

Carlson, M. Chalot, J. Chapman, G.-L. Chen, D. Cooper, P.M. Coutinho, J. Couturier, S.

Covert, Q. Cronk, R. Cunningham, J. Davis, S. Degroeve, A. Déjardin, C. Depamphilis,

J. Deter, B. Dirks, I. Dubchak, S. Duplessis, J. Ehlting, B. Ellis, K. Gendler, D.

Goodstein, M. Gribskov, J. Grimwood, A. Groover, L. Gunter, B. Hamberger, B. Heinze,

Y. Helariutta, B. Henrissat, D. Holligan, R. Holt, W. Huang, N. Islam-Faridi, S. Jones, M.

Jones-Rhoades, R. Jorgensen, C. Joshi, J. Kangasjärvi, J. Karlsson, C. Kelleher, R.

Kirkpatrick, M. Kirst, A. Kohler, U. Kalluri, F. Larimer, J. Leebens-Mack, J.-C. Leplé, P.

Locascio, Y. Lou, S. Lucas, F. Martin, B. Montanini, C. Napoli, D.R. Nelson, C. Nelson,

K. Nieminen, O. Nilsson, V. Pereda, G. Peter, R. Philippe, G. Pilate, A. Poliakov, J.

Razumovskaya, P. Richardson, C. Rinaldi, K. Ritland, P. Rouzé, D. Ryaboy, J. Schmutz,

J. Schrader, B. Segerman, H. Shin, A. Siddiqui, F. Sterky, A. Terry, C.-J. Tsai, E.

Uberbacher, P. Unneberg, J. Vahala, K. Wall, S. Wessler, G. Yang, T. Yin, C. Douglas, M.

Marra, G. Sandberg, Y. Van de Peer, D. Rokhsar (2006). The Genome of Black

Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. **313**: 1596-1604

Ushijima, K., R. Nakano, M. Bando, Y. Shigezane, K. Ikeda, Y. Namba, S. Kume, T. Kitabata, H.

Mori, Y. Kubo (2011). Isolation of the Floral Morph-Related Genes in Heterostylous Flax

(*Linum grandiflorum*): the Genetic Polymorphism and the Transcriptional and

- Post-Transcriptional Regulations of the S Locus. *The Plant Journal*. **69**: 317-331
- Vieira, C.P., D. Charlesworth, J. Vieira (2003). Evidence for Rare Recombination at the Gametophytic Self-Incompatibility Locus. *Heredity*. **91**: 262-267
- Vieira, J., E. Teles, R.A.M. Santos, and C.P. Vieira (2008). Recombination at Prunus S-Locus Region *SLFL1* Gene. *Genetics*. **180**: 483-491
- Webb, C.J. and D.G. Lloyd (1986). The avoidance of interference between the presentation of pollen and stigmas in angiosperms II. Herkogamy. *New Zealand Journal of Botany*. **24**: 163-178
- Xi, Z., B.R. Ruhfel, H. Schaefer, and C.C. Davis (2012). Phylogenomics and a Posteriori Data Partitioning Resolve the Cretaceous Angiosperm Radiation Malpighiales. *Proceedings of the National Academy of Sciences*. **109**: 17519-17524
- Yuan, S., G. Zeng, K. Zhang, S.C.H. Barrett (2023). Diverse Mating Consequences of the Evolutionary Breakdown of the Sexual Polymorphism Heterostyly. *Proceedings of the National Academy of Sciences*. **120**: e2214492120