

# An Agent-based Model of Pollen Competition in *Arabidopsis thaliana*

Charlotte Beckford\*, Montana Ferita<sup>†</sup>, Julie Fucarino<sup>‡</sup>  
Advisor: Alex Capaldi<sup>§</sup>

July 30, 2019

## Abstract

In 2016, Swanson et al. showed that when an *Arabidopsis thaliana* stigma is pollinated with equal amounts of pollen by two accessions, Columbia and Landsberg, Columbia pollen sire disproportionately more seeds. This phenomenon is known as nonrandom mating. Previous experiments have investigated nonrandom mating by examining how pollen performance traits such as proportion of pollen germinated, time to germination, and pollen tube growth rates differ between these two accessions. In addition, bioenergetics, such as the energy supplied to pollen tubes from the pistil during fertilization, likely also magnify competition. While plant fertilization is well-studied, the exact mechanics of pollen competition remain unknown. Using an agent-based model, we aim to identify the traits that cause pollen from one accession to sire more offspring than pollen from another accession and to what extent these traits contribute to this process. We calibrate our model against a number of parameters from empirical data to observe the output of seed siring proportions from mixed pollinations; we compare these values to those found in the literature. Our model can also be extended to predict seed siring proportions for other accessions of *Arabidopsis thaliana* given data on their pollen performance traits.

*Keywords:* agent-based modeling, pollen competition, non-random mating, mixed pollination, *Arabidopsis thaliana*

---

\*Fordham University, The Bronx, NY 10458

<sup>†</sup>Westminster College, New Wilmington, PA 16172

<sup>‡</sup>Wellesley College, Wellesley, MA 02481

<sup>§</sup>Valparaiso University, Valparaiso IN 46383

# 1 Introduction

Pollen competition refers to when accessions, or populations, of a plant species compete to fertilize ovules while using limited resources. Pollen competition has been observed in a variety of plant species. While this is a widely-observed phenomenon, little is known about the causes of pollen competition. In this paper we hope to identify which explanatory variables cause pollen competition and to quantify to what extent each variable contributes.

We are interested in pollen competition because of its impacts on pollination and, thus, agriculture. Gaining knowledge about pollen competition may contribute to agricultural efforts, an integral part of our society. In addition, we want to gain better understanding of a common but still fairly unknown phenomenon.

*Arabidopsis thaliana* is a well-studied plant: its entire genome has been sequenced and it serves as a model species for many other plants, including crops. This—coupled with the fact that when pollinated equally by two populations, one sires significantly more seeds than the other—makes *Arabidopsis thaliana* an ideal plant to study when modeling pollen competition [3]. Swanson et al. found that in mixed pollinations with the Columbia and Landsberg accessions Columbia pollen sired disproportionately more seeds than Landsberg with both Columbia and Landsberg pistils. This indicates that female choice in *Arabidopsis thaliana* does not play a significant role in one accession’s heightened performance over another in these mixed pollinations [4]. Thus, for simplicity, we only consider a Columbia *Arabidopsis thaliana* pistil being pollinated by the two accessions.

In 2016 Swanson et al. investigated potential causes of pollen competition when a Columbia *Arabidopsis thaliana* pistil is pollinated by Columbia and Landsberg accessions. Swanson et al. identified pollen performance traits that differ between these two accessions which could account for this disproportionate siring of seeds [9]. These traits include the proportion of pollen grains germinated, meaning that a greater proportion and, thus, more pollen grains of one accession may germinate than pollen grains of another; time to germination, as one accession may germinate faster than another accession; and pollen tube growth rate, as pollen tubes of one accession may grow faster than those of another accession. It was found that Columbia pollen has a greater proportion of pollen germinated than Landsberg, meaning that more total Columbia pollen grains germinate than Landsberg pollen grains. In addition, Columbia pollen also germinates faster than Landsberg pollen. Finally, Columbia pollen tubes grow faster than Landsberg pollen tubes.

Katherine Bassett previously developed an agent-based model simulating mixed pollination of an *Arabidopsis thaliana* pistil by two different accessions [1]. She focused on pollen tube growth patterns which may differ between the two accessions and potentially account for disproportionate seed siring. We are using her

work as a basis to develop a model which includes the pollen performance traits Swanson et al. found differed between Columbia and Landsberg accessions, as well as other aspects of pollination.

In addition to pollen performance traits found to differ between these two accessions, we also included bioenergetics and chemoattractants in our model. Bioenergetics refers to the ways that pollen tubes gain and use energy when growing. As pollen tubes of one accession may have more energy than those of another accession, whether because they may be able to gain more energy or because they require less energy to move and fertilize, bioenergetics may play a role in pollen competition. Chemoattractants emitted by unfertilized ovules control the direction of pollen tube growth [6]. When an ovule is fertilized, the polarity of the chemoattractants emitted by that ovule changes so the ovule begins emitting chemorepellents instead, preventing pollen tubes from growing towards that ovule. While many aspects of chemoattractants remains undiscovered, it is known that they grow stronger with time since germination and with proximity to the ovule. We consider chemoattractant strength in relation to different time points and distances in the pistil and how this may contribute to different seed siring proportions.

With these considerations in mind, our aim was to build an agent-based model representing this system, allowing us to investigate the role of different pollen performance traits in pollen competition. Agent-based models (ABMs) are a type of mathematical model that show how individuals interact with other individuals and with their local environment, called a world [8]. The world of an ABM is represented as a grid of patches characterized by properties which may differ between individual patches. As opposed to equation-based models, agents in an ABM are portrayed and interact as individuals. One of the advantages of agent-based models is that they can show causality in a system. While a number of processes are programmed directly into an ABM, other properties may emerge due to these programmed ones, establishing causality. Since we are hoping to identify and quantify which pollen performance traits cause pollen competition and to what extent each one contributes, an ABM was an appropriate choice for our model. Our model was built in NetLogo v6.1.0 [10], a modeling environment specifically designed for agent-based modeling. While a three-dimensional version of NetLogo exists, we chose not to utilize this version because it requires that the world wraps, meaning that all edges are connected to one another. Since our model is representing pollen tubes growing down a pistil, it is nonsensical for them to begin back at the top once they reach the bottom. While it is possible that physical occlusion plays a role in pollen competition, we do not consider it in this model. We believe that not addressing physical occlusion is reasonable because, while it is possible for pollen tubes to block one another while growing, it is also possible for pollen tubes to merely wind around one another while growing. In this way, other factors, namely the pollen performance traits, must account for one pollen accession outcompeting the other by fertilizing more ovules.

This paper is organized as follows: In Section 2, we will provide details of our agent-based model, using the standard Overview, Design Concepts, and Details (ODD) procedure [5]. In Section 3, we will discuss our results attempting to calibrate our agent-based model using distributions of pollen tube length. We conclude in Section 4 with a discussion of the results and suggestions for future work.

## 2 Methods

The Overview, Design Concepts, and Details (ODD) procedure is a protocol for describing ABMs [5]. The ODD procedure was developed as an attempt to facilitate recreate of simulation models by standardizing a protocol for describing them.

### 2.1 Assumptions

In our model we included various simplifying assumptions. In the literature, there exists a distinction between pollen viability and pollen germination. Pollen viability refers to the number of dead or aborted pollen initially placed on the stigma as part of the initial number of pollen. Pollen germination refers to the proportion of the initial number of pollen grains placed on the stigma which grow pollen tubes down into the ovary. However, for the purposes of this model the distinction between pollen viability and pollen germination proportion appeared redundant. The empirical data used to create this model provided pollen germination rates of the two accessions which accounted for the pollen viabilities of both accessions as well. Thus, we used the proportion of pollen germination in the setup of our model to encompass both variables of pollen viability and pollen germination.

When adding the time to germination ( $t_g$ ) variable to our model we assumed that the data fit a logistic curve. This suggestion was under the supervision of Capaldi. The data we used for time to germination is in Table 1. In addition to this data, we also included the data point (0,0) as part of the data set. At time zero, none of the pollen grains have had time to germinate; therefore, this is a reasonable assumption.

Pollen tube attrition is when pollen tubes stop growing in the style. While this is a known phenomenon, no data exists for what proportion of pollen tubes experience attrition. We assumed pollen tube attrition for both accessions was at least 50%, likely between 70% and 90%, based off of Swanson's estimate of pollen tube attrition from what he has observed in previous experiments.

We are assuming that the two accessions, Columbia and Landsberg, have the same sensitivity to chemoattractants released by the Columbia pistil. Chemoattractants help guide the pollen tubes towards unfertilized ovules. Also, we are

| Time (minutes) | Columbia  | Landsberg   |
|----------------|---|---|
| 10             | 17<br>75<br>0<br>0<br>17<br>8<br>6<br>2<br>3                | 0<br>3<br>4<br>4<br>2<br>0<br>1<br>0<br>0                   |
| 20             | 85<br>90<br>51<br>333<br>52<br>69<br>239<br>103<br>128      | 55<br>56<br>13<br>22<br>42<br>13<br>36<br>15<br>31          |
| 30             | 249<br>612<br>695<br>351<br>624<br>572<br>299<br>556<br>612 | 236<br>254<br>286<br>288<br>275<br>192<br>400<br>338<br>388 |
| 40             | 386<br>403<br>580<br>675<br>432<br>406<br>588<br>390<br>531 | 100<br>210<br>300<br>250<br>254<br>558<br>401<br>312<br>362 |

Table 1: Table of number of pollen grains germinated at various time points from biological experiments.

assuming that the radius of the chemoattractants overlaps for each neighboring ovule on the same side of the transmitting tract.

## 2.2 Purpose

It has been observed that in mixed pollination *Arabidopsis thaliana* displays non-random mating, meaning that when equal amounts of pollen from two unique accessions are placed on the stigma, one accession sires disproportionately more seeds than the other accession. Despite this phenomenon of pollen competition being widely observed, it remains unknown why it occurs. Previous investigation by Swanson et al. has analyzed pollen performance traits which differ between the Columbia or Col and Landsberg or Ler accessions which may account for Columbia siring disproportionately more seeds than Landsberg in mixed pollination. The purpose of this model is to gain a better understanding of pollen competition by determining which performance traits account for it and to what extent.

## 2.3 State variables and scales

There are three main classes of variables in our model, global variables, agent variables, and patch variables. In our model these agents are the Columbia and Landsberg pollen. The agents interact with each other and also with their environment or world which consists of a grid of patches. In our model, the world is the *Arabidopsis thaliana* pistil.

The global variables are shown in Table 2. In our model we represented the `stigma-length` as 10 patches and the `style-length` as 16 patches, where a square patch has a length of 0.0129mm. The variable `total-pollen` refers to the total number of pollen grains on the pistil which is randomly selected from a normal distribution with  $\mu = 1040$  and  $\sigma = 236$  [9]. The `prop-blue` variable refers to the proportion of the initial pollen grains which accounts for the Columbia accession. The `prop-red` variable has the same meaning except that it refers to the Landsberg accession. Thus, it is always the case that `prop-red` + `prop-blue` = 1. The variable `step-size` refers to the distance a single pollen tube can travel in one tick. We used a beta distribution to model the germination proportion for both of the accessions. We calculated the values of  $\alpha$  and  $\beta$  from Swanson's paper, and input these values as global variables. For the Columbia accession, `bluealpha` = 26.50 and `bluebeta` = 3.04 [9]. Likewise for the Landsberg accession, `redalpha` = 6.49 and `redbeta` = 2.80 [9]. Using the corresponding  $\alpha$  and  $\beta$  values for the two accessions, we are then able to determine the number of pollen grains that germinate as `blue-germ` and `red-germ`. `Blue-germ` is determined from the Beta Distribution using the corresponding `bluealpha` and `bluebeta` values, and likewise for `red-germ`. Each pollen tube picks a random number between 0 and 1, inclusive, and if the that number is

greater than **blue-germ** or **red-germ**, depending on the accession, then that pollen tube does not germinate. The number of pollen tubes from each accession that germinate at a given tick is represented as **B** for Columbia and **R** for Landsberg, which is modeled as Gompertz functions. Lastly, we defined **initial-tick** as the tick when the first pollen tube germinates.

Table 2: Global Variables

| Parameter            | Description   | Default Value | Source | Range  | Source |
|----------------------|---|---------------|--------|--|--------|
| <b>stigma-length</b> | Length of the stigma in patches                     | 10            |        | Not varied                                       |        |
| <b>style-length</b>  | Length of the style in patches                      | 16            |        | Not varied                                       |        |
| <b>total-pollen</b>  | Total number of pollen grains on the pistil         | N/A           |        | 1040 $\pm$ 236                                   | [9]    |
| <b>prop-blue</b>     | Proportion of Col pollen that pollinates the pistil | N/A           |        | 0 - 1  |        |
| <b>prop-red</b>      | Proportion of Ler pollen that pollinates the pistil | N/A           |        | 0 - 1  |        |
| <b>step-size</b>     | Distance a pollen tube can move in one tick         | N/A           |        | 0 - 2  |        |
| <b>bluealpha</b>     | Alpha value for Col germination proportion          | 26.50         | [9]    | Not varied                                       |        |
| <b>redalpha</b>      | Alpha value for Ler germination proportion          | 6.49          | [9]    | Not varied                                       |        |
| <b>bluebeta</b>      | Beta value for Col germination proportion           | 3.04          | [9]    | Not varied                                       |        |
| <b>redbeta</b>       | Beta value for Ler germination proportion           | 2.80          | [9]    | Not varied                                       |        |
| <b>blue-germ</b>     | Number of Col pollen grains that germinate          | N/A           |        | <b>prop-blue</b> $\times$ <b>total-pollen</b>    |        |
| <b>red-germ</b>      | Number of Ler pollen grains that germinate          | N/A           |        | <b>0 - prop-red</b> $\times$ <b>total-pollen</b> |        |
| <b>B</b>             | Number of Col pollen that germinates at each tick   | N/A           |        | 0 - 500  | [9]    |
| <b>R</b>             | Number of Ler pollen that germinates at each tick   | N/A           |        | 0 - 300  | [9]    |
| <b>initial-tick</b>  | Tick when first pollen tube germinates              | N/A           |        | 1 - 1440   |        |



The agent variables are shown in Table 3. The Columbia variables and the Landsberg variables are identical in their nature, but we allowed these variables to differ between the two accessions in order to see why one accession may outperform the other accession. In accordance with the research conducted by Swanson et al., we chose the color blue to represent the Columbia accession and the color red to represent the Landsberg accession. Thus, all Columbia variables are prefixed by **blue-** and all Landsberg variables are prefixed by **red-**. Without loss of generality, the following description of the variables for the Columbia accession also applies for the Landsberg accession. The variable **blue-initial-pollen-energy** is the amount of energy that all Columbia pollen tubes are assigned when they germinate. Pollen tubes need energy to move and fertilize, so this variable allows the pollen tubes to move after germinating. The variable **blue-energy-from-patch** refers to the amount of energy that a Columbia pollen tube is able to gain from a patch in one tick; the mechanism of this exchange is described in the Harvest Energy procedure. The **blue-fertilization-threshold** variable refers to the amount of energy that a Columbia pollen tube requires to fertilize an ovule. If a Columbia pollen tube is present on an unfertilized ovule but its energy does not meet the fertilization threshold, then the Columbia pollen tube is unable to fertilize the ovule at that tick. Lastly, there is **blue-movement-cost** which is the amount of energy required for the Columbia pollen tube to move forward a **step-size** at any tick. If the pollen tube does not have enough energy to move then it will remain in the same location until its energy meets **blue-movement-cost**.

Table 3: Agent Variables

| Parameter                                 | Description   | Default Value [Source] | Range [Source] |
|---|---|------------------------|----------------|
| <code>initial-blue-pollen-energy</code>   | Initial amount of energy of a pollen tube                             | N/A                    | 0 - 100        |
| <code>initial-red-pollen-energy</code>    | Initial amount of energy of a pollen tube                             | N/A                    | 0 - 100        |
| <code>blue-energy-from-patch</code>       | Amount of energy that a Col pollen tube can gain from a patch         | N/A                    | 0 - 10         |
| <code>red-energy-from-patch</code>        | Amount of energy that a Ler pollen tube can gain from a patch         | N/A                    | 0 - 10         |
| <code>blue-fertilization-threshold</code> | Amount of energy required for a Col pollen tube to fertilize an ovule | N/A                    | 0 - 20         |
| <code>red-fertilization-threshold</code>  | Amount of energy required for a Ler pollen tube to fertilize an ovule | N/A                    | 0 - 20         |
| <code>blue-movement-cost</code>           | Amount of energy required for a Col pollen tube to move               | N/A                    | 0 - 10         |
| <code>red-movement-cost</code>            | Amount of energy required for a Ler pollen tube to move               | N/A                    | 0 - 10         |

Apart from the agent variables there are patch variables which are shown in Table 4. There are two kinds of patch variables: Attractants and Bioenergetics. Attractants consists of the `chemoattractant-radius` and `slope` variables. All patches own an attractant value, which refers to the chemoattractant strength of a patch. This is the likelihood that a pollen tube will be attracted or, when the patch is emitting chemorepellents, repelled from it. The `chemoattractant-radius` defines the furthest distance from an ovule at which pollen tubes feel the effect of chemoattractants. The `slope` defines how quickly the chemoattractant strength grows to its maximum value. At time  $t = \frac{1}{\text{slope}} + \text{initial-tick}$ , the chemoattractants reach maximum strength. The other type of patch variable is bioenergetics. All patches have energy, which is called starch. All patches have `starch` except for patches that represent an ovule. The `initial-starch` variable refers to the initial energy of the patches. As pollen tubes grow down the transmitting tract they gain energy from the patches. This process of pollen tubes gaining energy and patches losing corresponding amounts of energy is described in the Harvest Energy procedure. Since the patches are losing energy in the model as pollen tubes grow down the transmitting tract, there is a function in the model that allows the observer to replenish the energy patches have. The options for replenishing energy are `no-replenish-starch`, `replenish-only-empty`, and `replenish-starch`. The `no-replenish-starch` option means that starch is never replenished in the model. There are two ways that starch can be replenished in the model, `replenish-only-empty` and `replenish-starch`. The `replenish-only-empty` procedure states that only patches which do not represent ovules and whose starch is completely depleted will replenish their starch. The `replenish-starch` option allows for all the patches that are not the ovules to replenish energy. Starch is replenished at specified times determined by `frequency-starch-replenish`. In the `replenish-only-empty` or `replenish-starch` procedure then the starch value of the patch is set to `proportion-starch-replenish` multiplied by `initial-starch`. The `replenish-starch` and `replenish-only-empty` procedures automatically turns off when the number of ovules fertilized is at least the `unfertilized-ovule-replenish-cutoff`.

Table 4: Patch Variables

| Parameter  | Description  | Default Value | Source | Range   | Source |
|--|--|---------------|--------|---------|--------|
| <code>chemoattractant-radius</code>              | Maximum distance pollen tubes can sense chemoattractants               | N/A           |        | 0 - 6   |        |
| <code>slope</code>                               | Determines how quickly chemoattractants reach maximum strength         | N/A           |        | 0 - 1   |        |
| <code>initial-starch</code>                      | How much starch non-ovule patches initially have                       | N/A           |        | 0 - 100 |        |
| <code>no-replenish-starch</code>                 | Replenish procedure does not occur                                     | N/A           |        | N/A     |        |
| <code>replenish-starch</code>                    | Replenish procedure occurs for all non-ovule patches                   | N/A           |        | N/A     |        |
| <code>replenish-only-empty</code>                | Replenish procedure occurs for all non-ovule patches whose starch is 0 | N/A           |        | N/A     |        |
| <code>unfertilized-ovule-replenish-cutoff</code> | Stops Replenish procedure  | 30            |        | 0 - 60  |        |
| <code>frequency-starch-replenish</code>          | Determines with what frequency patches replenish starch                | N/A           |        | 0 - 100 |        |
| <code>proportion-starch-replenish</code>         | Determines maximum amount of starch which a patch can replenish        | N/A           |        | 0 - 1   |        |

In addition to the state variables it is also important to include the scales present in the model. In NetLogo time is measured in ticks and space is measured in patches. Our model uses a scale of one tick = one minute and one patch =  $0.0129 \text{ mm} \times 0.0129 \text{ mm}$ . Our world consists of a 12 by 146 grid of patches and it represents the pistil of an *Arabidopsis thaliana* plant. The pistil is comprised of the stigma, style, and ovary.

The ratio of this grid is equivalent to the ratio of the height to the width of the pistil of an *Arabidopsis thaliana* plant (21.66:1). The size scale of the model was determined based off of the measurement of an *Arabidopsis thaliana* pistil and the number of patches which compose one in the model. The top 16 rows of patches represent the stigma, which is where the agents are initially placed. The next 10 patches represent the style. After the stigma and style is where the ovary starts. Research has found that the average number of ovules in an *Arabidopsis thaliana* ovary is between 50 and 70; we programmed 60 ovules into our model with 30 on each side of the ovary. The ovules have a size of two by three patches; however, before fertilization only one patch is indicated as the ovule. This single patch is the exact point that a pollen tube must grow to in order to fertilize an ovule. These ovules are equally spaced 4 patches apart. The agents represent the tips of pollen tubes; their descent down the ovary represents pollen tubes growing down the ovary in the transmitting tract towards the ovules.

When the pollen are created in the model, they are placed in a region representing the stigma. The length of this section was determined using an estimate of the ratio of stigma length to the rest of the pistil, which was about 1:8. Pollen in this region are not spread horizontally because the bottleneck effect that occurs in the style of *Arabidopsis thaliana* in vivo does not depend on the horizontal growth of the pollen tube as much as the vertical when occlusion is not being modeled.

Our next step was to give ticks a practical meaning. We chose to let ticks represent minutes, as some processes occur on a scale of minutes, making hours too large a measurement. Using the grant proposal, which states on page 501 that fertilization occurs within 8 to 10 hours, we determined that most or all of the ovules should be fertilized by the 600th tick, with some finishing around 480 ticks.

## 2.4 Process overview and scheduling

Each tick in the model represents one minute and each realization of the model runs for at most 24 hours. During each tick there are four pollen tube agent procedures, that occur in the following order: Germinate, Harvest Energy, Move, and Fertilize. It is possible for a pollen tube agent to not perform a process during a time step depending on if it meets the necessary conditions for that process; however, it is not possible for a pollen tube agent to perform these

processes out of this specified order. In addition, a fifth process, Replenish, which is a patch procedure, may occur if the user selects `replenish-starch` or `replenish-only-empty`. If this is the case, then during every tick divisible by `frequency-starch-replenish`, this process occurs after the four other processes which occur every tick. The decision tree represents all possible actions an individual pollen tube agent can perform during each tick is illustrated in Figure 1. It depicts the scheduling of these actions, which are the four processes repeated every tick. The start of the tick is depicted as a yellow circle, conditions which must be met before the processes occur are depicted as blue diamonds, processes are depicted as green circles, and stop points are depicted as red circles.

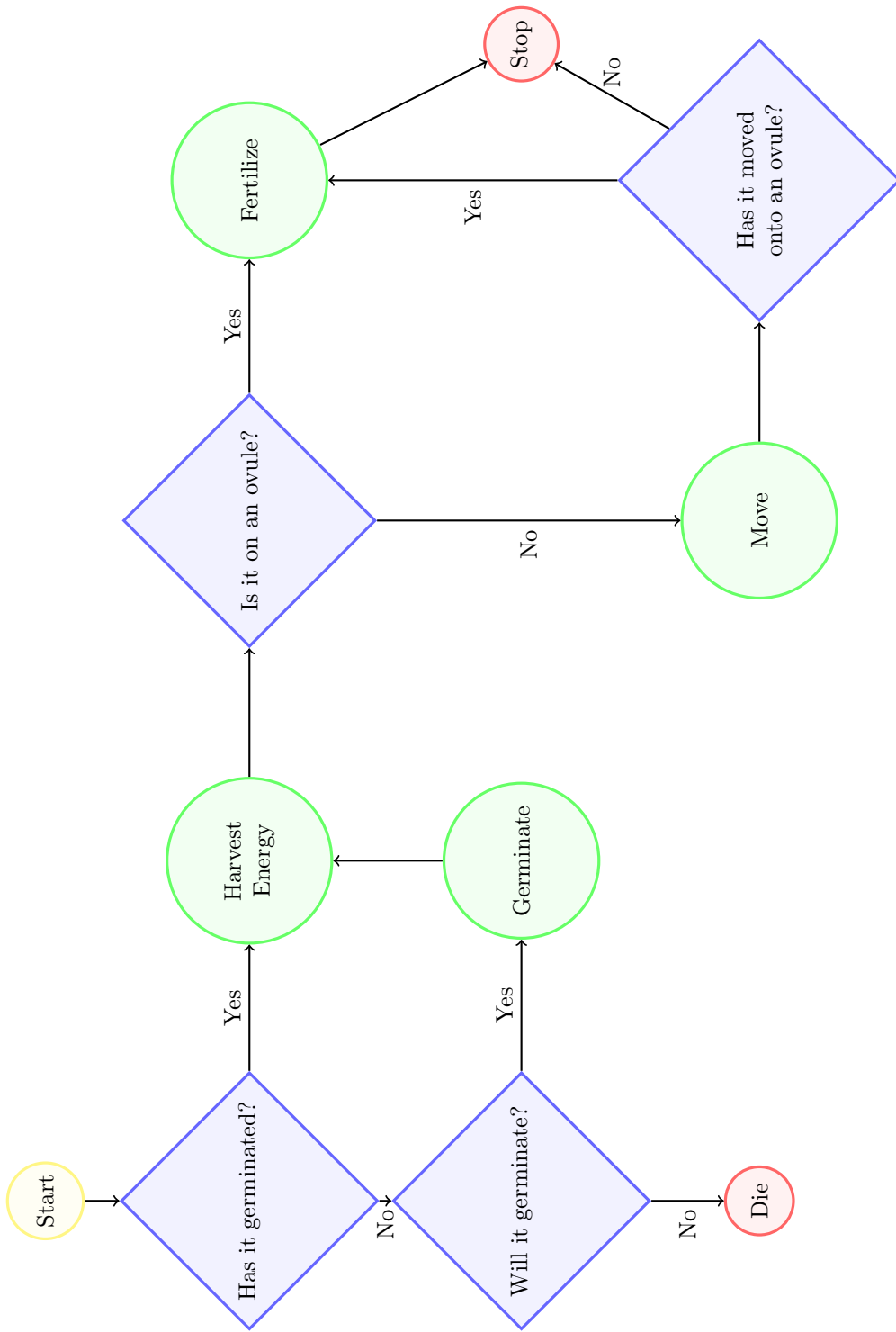


Figure 1: This figure shows all of the possible actions an individual pollen tube can perform each time step.

## 2.5 Design Concepts

### 2.5.1 Emergence:

The proportion of seeds sired by each accession emerges via competition between pollen tubes of different accessions to consume starch and reach unfertilized ovules. This is influenced by population-specific performance traits such as the proportion of pollen that will germinate and the time when pollen germinate, which we derived using experimental data. Pollen tube growth rate also emerges based on bioenergetics; however, without data on pollen energy or starch in the transmitting tract, the current values in the model do not have empirical bases.

### 2.5.2 Fitness:

At the beginning of each realization, the stigma is pollinated with 50% Columbia pollen and 50% Landsberg pollen. We know from experiments that *Arabidopsis thaliana* exhibits pollen competition, i.e. in mixed pollinations one accession sires disproportionately more seeds than the other. In Swanson et al.'s experiment Columbia sired more seeds than Landsberg. Therefore, fitness is measured by the proportion of seeds sired by each accession.

### 2.5.3 Sensing:

Pollen tubes detect chemical signals emitted from ovules which guide pollen tube growth. Unfertilized ovules emit chemoattractants and fertilized ovules emit chemorepellents. The chemical signal strength is a function of time since pollen tubes begin germinating and of distance from the closest ovule. The strength is factored into a weighted probability that the pollen tube will respond to the chemical, either by facing towards a chemoattractant or by facing away from a chemorepellent.

### 2.5.4 Interaction:

Pollen tubes interact indirectly with other agents by competing for limited starch and unfertilized ovules. While this competition occurs on an individual level, since pollen of the same accession share pollen performance traits, one accession may outcompete the other, which would manifest as one accession fertilizing over 50% of all fertilized ovules.

### 2.5.5 Stochasticity:

After the pollen grains have been placed on the stigma, the proportion of pollen that will germinate from each accession is drawn randomly from a Beta distri-



bution derived from experimental data specific to that accession. In particular, our data was derived from Swanson et al.’s experiment. While the number of pollen from each accession that will germinate at a given tick  $t$  is modeled with a deterministic equation, the particular agents that germinate at that time are chosen randomly from their respective populations.

Additionally, the Move procedure incorporates stochasticity at several points. When pollen tubes are not in the `chemoattractant-radius` of an ovule, they will move in a random downward walk, as we know from experiments that pollen tubes are only capable of growing downwards in the absence of chemical signals. When a pollen tube can detect chemical signals, the intensity of these, ranging from zero to one, is compared to a randomly-generated decimal between zero and one (not inclusive of one). If the chemical signal strength is greater than the randomly-generated value, the pollen tube will react to it. If the chemical signal is being released by an unfertilized ovule, meaning it is a chemoattractant, the pollen tube will face towards the ovule. If the chemical signal is being emitted by a fertilized ovule, meaning it is a chemorepellent, the pollen tube will face away from the ovule.

### 2.5.6 Collectives:

When pollen are created, they are assigned one of two accessions, Columbia or Landsberg. The accession determines the distribution of the pollen germination proportion and the number of pollen that germinate at each tick, which influence fertilization patterns. Additionally, the user may vary bioenergetic parameters by accession—within an accession, these traits are the same.

### 2.5.7 Observation:

Previous experiments with *Arabidopsis thaliana* pollen tubes grown in vivo have demonstrated that for both populations, at time intervals when pollen tube lengths were measured, the distributions of pollen tube lengths does not differ significantly from an exponential distribution. Since we let one minute correspond to one tick, we were able to measure the lengths of all pollen tubes present at the corresponding time in ticks and compare the distribution of these lengths to the theoretical distribution for that accession at that time.

## 2.6 Initialization

At the beginning of a run the agents and environment are initialized to certain values, some of the which are determined by the observer while others are randomly chosen from a distribution. The number of ovules are set to 60, with 30 ovules on the left side and 30 ovules on the right side. The initial values of the stated variables are determined by the observer. The only stated variable

that is visually displayed in the model is the `chemoattractant-radius`. In the model, the `chemoattractant-radius` is represented by green patches. A larger `chemoattractant-radius` value corresponding to a greater area of the transmitting tract being shown in green. Also a change in the value of the `chemoattractant-radius` affects the location of the pollen grains. A larger `chemoattractant-radius` corresponds to a smaller width where the pollen grains can be placed before the beginning of the run. Swanson noticed in his experiments that the first couple of ovules in the beginning of the transmitting tract tend to never get fertilized. When we did not account for the dependence of the width of the pollen grains location on the `chemoattractant radius`, the ovules at the top of the transmitting tract almost certainly were fertilized. After the pollen grains formed pollen tubes, the pollen tubes that germinated first and or within the range of the `chemoattractant-radius` immediately fertilized the first few ovules.

In Swanson et al.'s experiment the total amount of pollen placed on the stigma was  $1040 \pm 236$  pollen. Therefore, in our model we chose the total number of pollen to be randomly chosen from a Normal distribution with a  $\mu = 1040$  and a  $\sigma = 236$ . In order to determine how many of the total pollen grains are Columbia or Landsberg, we allow the user to adjust the value of the proportion of the Landsberg pollen of the total number of pollen represented as `prop-red`. Thus the total amount of Landsberg pollen is equivalent to `total-pollen`  $\times$  `prop-red`. Similarly, the total number of Columbia pollen is equal to `total-pollen`  $\times$  `prop-blue` where `prop-blue` =  $1 - \text{prop-red}$ . Therefore, it is now possible for Columbia and Landsberg to have different initial numbers of pollen grains before germination.

Germination occurs after pollen attach to a stigma. During germination, pollen grains are hydrated by the pistil and a pollen tube will grow through a special hole in the wall of the pollen grain, called an aperture, before growing down the style towards the transmitting tract. In our model the number of pollen grains placed on the stigma is assumed to be equal for both accessions. However, accessions have different pollen germination proportions, meaning that even when equal numbers of pollen grains of two accession are placed on a pistil, one accession will grow more pollen tubes than the other. The pollen which do not germinate include both nonviable pollen, meaning pollen grains which are incapable of germinating because they are damaged for some reason, and also those which simply do not germinate.

From Swanson et al. we know that the proportion of pollen germinated for Columbia was 89.7% while for Landsberg it was 69.9%. While Swanson et al. did not provide a percent error for these measurements, we were able to calculate these values based off of absolute errors the paper included for the numbers of ungerminated pollen. The paper indicates that, during single-donor pollinations of an *Arabidopsis thaliana* pistil, there were  $107 \pm 56.9$  ungerminated Columbia pollen grains and  $313 \pm 149.1$  ungerminated Landsberg pollen grains. This means that 107 pollen grains is correspondent to  $100\% - 89.7\% = 10.3\%$  of the

total Columbia pollen and that 313 pollen grains is correspondent to 100% - 69.9% = 30.1% of the total Landsberg pollen.

Using this information, we were able to calculate the total number of pollen grains for each accession. Rounding to three significant figures, there were in total 1040 pollen grains for both Columbia and Landsberg accessions. Dividing the absolute uncertainty by the total number of pollen grains and multiplying by 100 we approximated the percent error of pollen germination proportion to be 5.48% for Columbia and 14.3% for Landsberg.

We used these values to estimate  $\mu$  and  $\sigma$  for the proportion of Columbia pollen germinated ( $\hat{p}_C$ ) and proportion of Landsberg pollen germinated ( $\hat{p}_L$ ), with  $\hat{p}_C = 0.897 \pm 0.0548$  and  $\hat{p}_L = 0.699 \pm 0.143$ . We decided to use a Beta distribution to model the proportion of pollen that germinate, as it is a distribution of values from zero to one, making it useful for proportions. We found  $\alpha$  and  $\beta$ , the parameters for a Beta distribution, for each accession, using the following equations:

$$\alpha = \mu \left( \frac{\mu(1-\mu)}{\sigma^2} \right) \quad (1)$$

$$\beta = (1-\mu) \left( \frac{\mu(1-\mu)}{\sigma^2} \right), \quad (2)$$

where  $\mu$  is the proportion of pollen germinated from each accession and  $\sigma$  is the percent error. First we had to verify that  $\sigma^2 < \mu(1-\mu)$  for our estimations of  $\mu$  and  $\sigma$  for each accession, but this condition was met, so we were able to compute the parameters for the Beta distributions of the proportion of pollen germinated for both accessions. For Columbia, the parameters were  $\alpha=26.5$  and  $\beta=3.04$ , while for Landsberg they were  $\alpha=6.49$  and  $\beta=2.80$ . These  $\alpha$  and  $\beta$  values are defined as global parameters for each accession in the model. At the beginning of each realization, the model randomly selects a value from the Beta distribution corresponding to each accession; this represents the proportion that should germinate from the population during the current realization. Next, each pollen tube agent randomly selects a number from zero to one, non-inclusive. If the randomly-generated number is greater than the associated germination proportion for that realization, it dies. Otherwise, the pollen tube germinates. This determines the total number of pollen grains of each accession that germinate and ensures it matches what we know from empirical data.

## 2.7 Input

Chemoattractants in the model follow a Holling Type I saturating functional response curve. They first appear after germination begins, i.e. after the first pollen germinates, and after that point chemoattractant strength increases over time until a critical point is reached. This point is determined by the user, as there is currently a lack of data on exactly how chemoattractant strength changes over time in *Arabidopsis thaliana*.

Chemoattractant strength also depends on how close a pollen tube is to an ovule. Once pollen tubes are within the predetermined `chemoattractant-radius` of an ovule, the pollen tube will be more attracted, i.e. more likely to move towards the ovule, as it gets closer. The `chemoattractant-radius`, which is constant for both populations, is user-controlled, as it is unknown how chemoattractant strength varies by location in the pistil.

## 2.8 Submodels

### 2.8.1 Germinate:

The germinate process accounts for differential pollen tube germination rates. Some accessions germinate faster than others, meaning that pollen tubes of that accession may be able to fertilize ovules before pollen tubes of the other accession, providing a potential advantage for fertilizing ovules. From Swanson et al. we obtained values for the average number of pollen grains germinated at various time points and modeled a Gompertz function, a generalized form of a Logistic curve of the form

$$N(t) = N_0 \exp\left(\frac{r_0}{\alpha}(1 - \exp(-t\alpha))\right), \quad (3)$$

where  $r_0$  is equivalent to  $k$ , the growth capacity,  $N_0$  is the first data point, and the carrying capacity  $L = N_0 \exp\left(\frac{r_0}{\alpha}\right)$ , to the data. For Columbia the calculated value of  $\alpha = \frac{r_0}{3.5600}$  and for Landsberg the calculated value of  $\alpha = \frac{r_0}{5.2619}$ . When we programmed these equations into MATLAB[7], with  $k$  or  $r_0$  being the one parameter being fitted, MATLAB estimated that the  $k$  value for Columbia was 0.2931 and for Landsberg was 0.4811. Thus, the model uses the equation

$$N_C(t) = 14.22 \exp(3.560(1 - \exp(-0.08233 * t))) \quad (4)$$

to model the number of Columbia pollen tubes which germinate per minute, meaning that they can then move down the transmitting tract, and

$$N_L(t) = 1.5556 \exp(5.2619(1 - \exp(-0.09143 * t))) \quad (5)$$

to model the same thing for Landsberg. In our model, all pollen tubes are initially ungerminated, so all of them have a property called `germinate` which is set to false. During the Germinate procedure a number equivalent to the value of the functions at that tick will germinate for each accession. Once germinated, the pollen tube's property `germinate` becomes true. The pollen tube then sets its energy to the initial pollen energy parameter: `initial-blue-pollen-energy` for Columbia and `initial-red-pollen-energy` for Landsberg.

### 2.8.2 Harvest Energy:

After pollen tubes have germinated, they perform the Harvest Energy procedure during each subsequent time step. First they check to see what the starch value of the patch they are currently located on is. Each accession has a corresponding parameter which determines how much starch they may harvest from a patch: for Columbia the parameter is `blue-energy-from-patch` and for Landsberg the parameter is `red-energy-from-patch`. If the patch that a pollen tube is currently located on has starch greater than or equal to the parameter that determines how much starch that pollen tube receives from a patch, then the pollen tube gains energy equivalent to that parameter value and the patch loses starch equivalent to that value. Otherwise, the pollen tube does not harvest any energy.

### 2.8.3 Move:

Once a pollen tube is germinated, it may perform the Move procedure. The direction of pollen tube growth is controlled by chemical signals emitted by unfertilized ovules. Chemical signals turn on as soon as the first pollen tube germinates in order to guide pollen tubes to grow towards unfertilized ovules. The purpose of chemoattractants is to guide the growth of pollen tubes towards an unfertilized ovule. When an ovule is fertilized, it stops emitting chemoattractants and emits chemorepellents. The role of chemorepellents is to ensure that exactly one pollen tube fertilizes each ovule. Chemoattractant strength is relative to both time since germination and distance from ovule. For parsimony, we created a Holling's type I saturating functional response curve to represent the chemoattractant strength. This means that the chemoattractant strength increases linearly until it reaches its maximum value; once it reaches its maximum value it stays constant. To simplify coding, we limited the values of chemoattractant strength to a range from 0 to 1, inclusive. Then the chemoattractant strength can represent a probability that a pollen tube will be attracted to an ovule. Since chemoattractant strength is relative to both time and distance it can be represented by the general formula

$$A_i(x, y, t) = F(t) \left( \frac{d_{max} - d(x, y)}{d_{max}} \right), \quad (6)$$

where  $d_{max}$  is a variable slider which we call `chemoattractant-radius` and it represents the maximum distance from an unfertilized ovule at which a pollen tube can sense a chemoattractant.  $F(t)$  represents the time component of this function, so it reflects how chemoattractant strength increases relative to time since germination. In our code,  $F(t)$  is referred to as the attractant multiplier. We defined the time when the first pollen grain has germinated as the `initial-tick`. The number of ticks it takes for the chemoattractants to reach their maximum strength is determined by a slider parameter called `slope`. The

reciprocal of the slope signifies how many ticks it takes for the chemoattractants of an unfertilized ovule to increase to 1. Thus, when the number of ticks since the first pollen germinated, `ticks - initial-tick`, is less than the number of ticks it takes for the chemoattractant to reach its maximum strength, `1/slope`,  $F(t)$  is defined as `slope(ticks - initial-tick)`. Once the tick difference is greater than the reciprocal of the slope, the attractant multiplier is set as 1. Thus, after this point, the attractant strength varies only according to distance, not time. For simplicity, rather than programming chemorepellents as a different function than chemoattractants, when an ovule is fertilized the chemoattractant strength now represents the chemorepellent strength. The value is the same, but the response of the pollen tube now differs.

The procedure is as follows and is illustrated in Figure 2. First pollen tubes check to see if there is an unfertilized ovule in its `chemoattractant-radius`, which is the maximum distance from an ovule where chemoattractants can be sensed by pollen tubes. If an unfertilized ovule is present in a pollen tube's `chemoattractant-radius`, then it tests to see if it is attracted. It does this by randomly generating a number from 0 to 1, noninclusive. If the number generated is less than the maximum attractant value of a patch in the pollen tube's `chemoattractant-radius`, then the pollen tube is attracted. It will face a patch with that maximum attractant value. If no unfertilized ovules are present in a pollen tube's `chemoattractant-radius`, it checks to see if there are any fertilized ovules present. If there are fertilized ovules but no unfertilized ovules in its `chemoattractant-radius` then the pollen tube checks to see if it repelled. Again, a number is generated randomly from 0 to 1 noninclusive. If the number is less than the maximum attractant value, which now represents a repellent value as the attractant values in this scenario are all being emitted by fertilized ovules, then the pollen tube is repelled. It faces the opposite x-direction of a patch with the maximum attractant value, as well as a random y-coordinate below its current position. Finally, if there are neither fertilized nor unfertilized ovules present in the `chemoattractant-radius`, then the pollen tube faces a random x-coordinate and a random y-coordinate below its current position. Once the pollen tube is facing the appropriate direction, if it has enough energy to move then it will move forward a `step-size`. The amount of energy a pollen tube requires to move is determined by the parameters `blue-movement-cost` and `red-movement-cost`, for Columbia and Landsberg, respectively. While the energy required for a pollen tube to move varies between the accessions, for parsimony it is assumed that all individuals in an accession require the same amount of energy to move. When a pollen tube moves forward a `step-size`, it loses energy correspondent to the amount of energy it requires to move. If the pollen tube does not have enough energy to move then it just stays on the patch it is on, while still facing the appropriate direction.

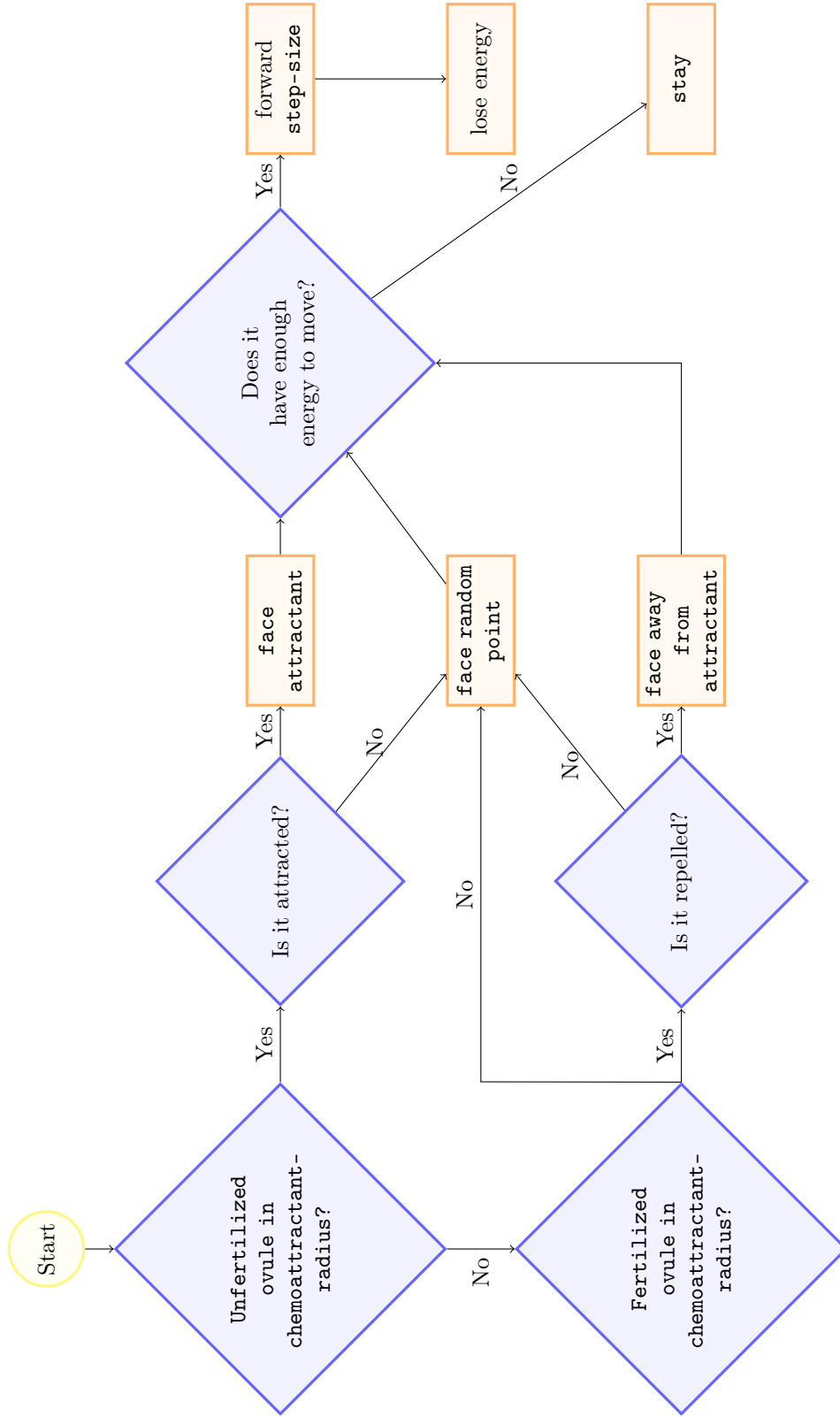


Figure 2: A flow chart of the move procedure.

#### 2.8.4 Fertilize:

The fertilize procedure occurs only when a pollen tube is located on a patch which is an unfertilized ovule. If the pollen tube has enough energy to fertilize the ovule, then it does so. The amount of energy a pollen tube needs to fertilize an ovule varies between the two accessions; it is defined as `blue-fertilization-threshold` for Columbia and `red-fertilization-threshold` for Landsberg. When a pollen tube fertilizes an ovule it changes the color of the ovule and of the ovule's neighbors to the color corresponding to the pollen tube's accession—blue for Columbia and red for Landsberg. The reason why the ovule's neighbors change color is because all of these patches represent the entire ovule while before the single patch which represented the unfertilized ovule actually represented the exact location on an ovule that a pollen tube must grow to in order to fertilize it. Afterwards the pollen tube dies. Now that the ovule is fertilized, instead of emitting chemoattractants it will emit chemorepellents.

#### 2.8.5 Replenish:

The Replenish procedure only occurs when the observer selects `replenish-starch` or `replenish-only-empty` and when the number of unfertilized ovules is no greater than `unfertilized-ovule-replenish-cutoff`. When this is the case, the procedure occurs at ticks divisible by the parameter `frequency-starch-replenish`. If the user selects the `replenish-starch` option, then at these time steps the starch of all patches which do not represent an ovule and which, if they gain the starch available will still have `starch` less than the `initial-starch`, then `starch` is set to the value `starch + (proportion-starch-replenish) (initial-starch)`. If gaining the available starch will make the starch greater than or equal to `initial-starch`, then starch is just set to `initial-starch`. If the `replenish-only-empty` option is chosen, then, at these time steps, the starch of patches with no starch will be set to `(proportion-starch-replenish) (initial-starch)`. Thus, the starch is replenished after being completely depleted by pollen tubes gaining energy. In both cases, when the number of unfertilized ovules is less than the `unfertilized-ovule-replenish-cutoff`, the Replenish procedure stops occurring. This is a parsimonious way to represent diminishing returns.

## 3 Results of Model Calibration

### 3.1 Number of pollen grains germinated:

We used literature values to create a sigmoidal function for both accessions to model the number of pollen grains germinated in relation to time. We used MATLAB to compute logistic and Gompertz models varying one, two, and



three parameters for the Columbia pollen to find the function which best fit the model. We used the standard form of a logistic function

$$y = \frac{\mathbb{L}}{1 + A \exp(-kx)}, \quad (7)$$

where  $\mathbb{L}$  is the carrying capacity,  $k$  is the growth rate, and  $A$  is a coefficient. When assuming that these were constant, we assumed that  $\mathbb{L}$  was 500, based off of projecting a graph depicted in Swanson et al., that  $k$  was 0.01, and that  $A$  was 1. We also computed the *AICc* to measure which model was most accurate while using the least number of parameters; this formula is meant for smaller data sets, which our data set qualifies as since we only had 37 data points per accession to model the pollen germination. Smaller *AICc* values indicate that a model is more appropriate for a data set. The equation to calculate *AICc* is

$$AICc = 2k - 2\ln(\hat{L}) + \frac{2k(k+1)}{n-k-1}, \quad (8)$$

where  $k$  is one more than the number of fitted parameters,  $\hat{L}$  is the likelihood function, and  $n$  is the number of data points. One is added to the number of fitted parameters to account for variance. The likelihood function  $\hat{L}$  is measured as

$$y = \hat{L} = \frac{SSE}{n}, \quad (9)$$

where  $n$  is the number of data points.

| Function | $N$ | Parameters         | AICc   | $\Delta$ |
|----------|-----|--------------------|--------|----------|
| Gompertz | 1   | $k$                | -15.58 | 1.82     |
| Gompertz | 3   | $A, k, \mathbb{L}$ | -9.74  | 7.66     |
| Logistic | 1   | $A$                | -17.40 | 0.00     |
| Logistic | 1   | $k$                | -17.19 | 0.21     |
| Logistic | 1   | $\mathbb{L}$       | -17.40 | 0.00     |
| Logistic | 2   | $A, k$             | -11.78 | 5.62     |
| Logistic | 2   | $A, \mathbb{L}$    | -14.74 | 2.66     |
| Logistic | 2   | $k, \mathbb{L}$    | -14.64 | 2.76     |
| Logistic | 3   | $A, k, \mathbb{L}$ | -9.26  | 8.14     |

Table 5: Table of the various fits for the model along with their corresponding *AICc* values.

The list of the *AICc* values we calculated for the various fits is shown in Table 5, where  $N$  is the number of fitted parameters. We found that, while the logistic models which varied only one parameter had the lowest *AICc* values, indicating that they are most appropriate to model the data, the graphs of these functions actually did not seem to model the data well. While Swanson's data indicated that the 0 time minutes could be used in the data for pollen germination per

time, which makes sense since at  $t = 0$  minutes 0 pollen grains would be germinated, we did not use 0 as the initial data point. Instead, for the functions for both Columbia and Landsberg accessions we used the average number of pollen grains at  $t = 10$  minutes as the initial data points. Based off of the graphs in Swanson’s report which corresponded to the measurements of pollen grains over time, we estimated that the carrying capacity for Columbia was 500 and for Landsberg 300.

### 3.2 Distribution of pollen tube length:

We tested our model against proposed distributions of pollen tube lengths for each accession at four time points post-pollination.

Previous research has shown that the distributions of Columbia and Landsberg pollen tube lengths after three, six, nine, and twenty-four hours did not differ significantly from the data from Swanson et al. [2].

We ran an experiment varying eight parameters for a total of 2304 parameter combinations, which are listed in Table 6. In the experiment each parameter combination had 10 realizations. For each realization, we tested the distributions of pollen tubes of each accession at each time against those proposed by Capaldi and Kolba by running Kolmogorov-Smirnov (K-S) two-sample tests at the 0.05 significance level. A two-sample K-S test is used to determine if two samples of data come from the same distribution. For each K-S test, we generated a random sample of numbers from the exponential distributions modeled by Capaldi and Kolba and tested to see if the model data and corresponding random sample came from the same distribution: that is, if the model data came from the same exponential distribution.

When we ran K-S tests for each realization of the model data against the exponential distributions, exactly one realization of each accession failed to reject the null hypothesis for two time points, 3 hours and 6 hours. For the Columbia accession the corresponding parameter combination was `step-size 0.50, slope 0.500, chemoattractant-radius 4, blue-energy-from-patch 0.50, red-energy-from-patch 0.50, blue-movement-cost 2.00, red-movement-cost 0.50`, and replenish set to `no-replenish-starch`. For the Landsberg accession the corresponding parameter combination was `step-size 0.50, slope 0.050, chemoattractant-radius 5, blue-energy-from-patch 2.00, red-energy-from-patch 2.00, blue-movement-cost 2.00, red-movement-cost 2.00`, and replenish set to `no-replenish-starch`.

To further analyze the data, we ran two-sample K-S tests for every realization against gamma distributions Capaldi and Kolba modelled to the pollen tube length distributions of each accession at the four time points. For the Columbia accession only one realization failed to reject the null hypothesis for two time points, 9 hours and 24 hours. This parameter combination was `step-size 2.00`,

| Parameter Combinations |   |
|------------------------|---|
| Variable               | Values  |
| step-size              | 0.50 1.00 1.50 2.00   |
| slope                  | 0.005 0.050 0.500   |
| chemoattractant-radius | 2 3 4 5   |
| blue-energy-from-patch | 0.50 2.00   |
| red-energy-from-patch  | 0.50 2.00   |
| blue-movement-cost     | 0.50 2.00   |
| red-movement-cost      | 0.50 2.00   |
| replenish              | replenish-starch<br>replenish-only-empty<br>no-replenish-starch |

Table 6: All of the parameters varied in our experiment to attempt to calibrate the model.

slope 0.500, chemoattractant-radius 4, blue-energy-from-patch 0.50, red-energy-from-patch 2.00, blue-movement-cost 2.00, red-movement-cost 2.00, and replenish set to no-replenish-starch. Forty-eight parameter combinations had at least one realization which failed to reject at a single time point. In fact, for a single parameter combination multiple realizations would fail to reject at a single time point or, in one case, at two time points. Three parameter combinations failed to reject different realizations at 3 hours and 6 hours. The parameter combination step-size 1.00, slope 0.050, chemoattractant-radius 3, blue-energy-from-patch 0.50, red-energy-from-patch 2.00, blue-movement-cost 2.00, red-movement-cost 0.50, and replenish set to no-replenish-starch failed to reject at 3 hours for one realization and at 6 hours for three different realizations. The parameter combination step-size 1.00, slope 0.050, chemoattractant-radius 5, blue-energy-from-patch 0.50, red-energy-from-patch 2.00, blue-movement-cost 2.00, red-movement-cost 0.50, and replenish set to no-replenish-starch failed to reject at 3 hours for two realizations and at 6 hours for one realization. Finally, the parameter combination step-size 1.00, slope 0.500, chemoattractant-radius 4, blue-energy-from-patch 0.50, red-energy-from-patch 2.00, blue-movement-cost 2.00, red-movement-cost 2.00, and replenish set to no-replenish-starch failed to reject at 3 hours for one realization and at 6 hours for four different realizations. Notably, this parameter combination had the most realizations that failed to reject at a single time point, four, and was the only parameter combination which had this many realizations that failed to reject at a single time point.

For the Landsberg accession not a single realization failed to reject the null hypothesis that the model data and the random sample of numbers generated from the gamma distribution at the corresponding time point and accession come from the same distribution at the 0.05 significance level.

## 4 Discussion and Future Work

In order to improve the model by eliminating the assumption of the equal sensitivities of the accessions to the chemoattractants, the following experiment can be conducted. It has already been noted that placing pollen grains on a plate with agar growth medium leads to pollen tubes growing. In order to test the sensitivity of Columbia and Landsberg accessions to Columbia chemoattractants, we can place pollen grains from both accessions on a plate with agar growth medium along with Columbia ovules and observe the growth direction of the pollen tubes from both accessions. If one accession is more sensitive to the chemoattractants than the other accession, then the pollen tubes will veer towards the ovules earlier on than those of the other accession.

It is interesting to note that both of the parameter combinations that failed to reject the null hypothesis that they follow the Columbia exponential distributions modelled by Capaldi and Kolba have small `step-size` values, large values for `chemoattractant-radius`, the movement cost of the accession for which the parameter combination failed to reject was set to 2.00, and `replenish` was set to `no-replenish-starch`. The value of `chemoattractant-radius` also impacts where the pollen are placed on the stigma initially, which may lead to more competition for starch if the pollen grains are grouped closer together. However, since only one out of ten realizations for both of these parameter combinations resulted in this fail to reject the null hypothesis at two time points, it is possible that these results are merely due to random chance.

Further analysis supported this. The gamma distribution is a generalization of the exponential distribution. Thus, the gamma distribution should have better modeled the distributions of pollen tube lengths. Considering this, and that there still were not promising results of parameter combinations which consistently failed to reject the null hypothesis, computing K-S tests for the gamma distributions further suggests that the realizations which failed to reject the null hypothesis may have done so merely because of random chance.

The results of the K-S test for the gamma distributions were slightly more promising for Columbia, but worse for Landsberg. While only one realization for Columbia failed to reject the null hypothesis that it did not come from the gamma distribution for Columbia pollen tube length modelled by Capaldi and Kolba, which is no better than the K-S test results for the exponential distributions of Columbia pollen tube length, multiple parameter combinations had different realizations which failed to reject at two time points. The Columbia pollen tube lengths from these realizations were better modelled by the gamma distributions. Not a single realization failed to reject the null hypothesis that Landsberg did not come from the gamma distributions modelled by Capaldi and Kolba. This suggests that the single realization which failed to reject that a model realization came from an exponential distribution for Landsberg pollen tube length was due to random chance.

In future experiments it is necessary to vary a wider range of parameter combinations. At the same time, it might be beneficial to slightly vary the values which are found similar between these two parameter combinations to determine if a better value for each exists. Hopefully future work can find ranges of parameter combinations which result in distributions of pollen tube lengths that correspond to those seen in experiments. Then the model will be calibrated against distributions of pollen tube lengths. Further experiments determining which of these parameter combinations result in the disproportional siring of seeds by the two accessions seen in experiments will narrow down possible values of parameters. This can help to determine and quantify what variables are responsible for this phenomenon, our central question. Finally, while we designed the model using data of the Columbia and Landsberg accessions and would like to recreate the results of mixed-pollination by these accessions, we would also like our model to be able to predict the seed siring proportions when *Arabidopsis thaliana* is pollinated by other accessions. Using data about these other accessions the model can be calibrated to these pollinations and thus will be able to predict the outcomes of different mixed-pollinations.

## 5 Acknowledgments

We would like to thank the Valparaiso Experience in Research by Undergraduate Mathematicians (VERUM) program and our advisor, Alex Capaldi. We would also like to thank our graduate assistant, David Elzinga, for his various contributions, and our biology advisor, Robert Swanson, and Tiffany Kolba for helping us with statistics. Finally, we thank the National Science Foundation for funding our research under grants DMS-1559912 and IOS-1645508.

## References

- [1] Katherine Bassett. Agent-based modeling of pollen competition. 2018.
- [2] Alex Capaldi and Tiffany N Kolba. Using the sample maximum to estimate the parameters of the underlying distribution. *PloS one*, 14(4):e0215529, 2019.
- [3] Ann L Carlson, Megan Telligman, and Robert J Swanson. Incidence and post-pollination mechanisms of nonrandom mating in *arabidopsis thaliana*. *Sexual plant reproduction*, 22(4):257–262, 2009.
- [4] Ann Louise Carlson, Jonathan Nesbit Fitz Gerald, Megan Telligman, Jacob Roshanmanesh, and Robert John Swanson. Defining the genetic architecture underlying female-and male-mediated nonrandom mating and seed yield traits in *arabidopsis*. *Plant physiology*, 157(4):1956–1964, 2011.

- [5] Volker Grimm, Uta Berger, Finn Bastiansen, Sigrunn Eliassen, Vincent Ginot, Jarl Giske, John Goss-Custard, Tamara Grand, Simone K Heinz, Geir Huse, et al. A standard protocol for describing individual-based and agent-based models. *Ecological modelling*, 198(1-2):115–126, 2006.
- [6] Tetsuya Higashiyama, Haruko Kuroiwa, Shigeyuki Kawano, and Tsuneyoshi Kuroiwa. Guidance in vitro of the pollen tube to the naked embryo sac of *torenia fournieri*. *The Plant Cell*, 10(12):2019–2031, 1998.
- [7] MATLAB. *version 9.6.0 (R2019a)*. The MathWorks Inc., Natick, Massachusetts, 2019.
- [8] Steven F Railsback and Volker Grimm. *Agent-based and individual-based modeling: a practical introduction*. Princeton university press, 2019.
- [9] Robert J Swanson, Adam T Hammond, Ann L Carlson, Hui Gong, and Thad K Donovan. Pollen performance traits reveal prezygotic nonrandom mating and interference competition in *arabidopsis thaliana*. *American Journal of Botany*, 103(3):498–513, 2016.
- [10] Uri Wilensky. NetLogo: Center for connected learning and computer-based modeling. *Northwestern University, Evanston, IL*, 1999.