

# SSWD-EvoEpi

## A Coupled Eco-Evolutionary Epidemiological Model for Sea Star Wasting Disease

---

**Willem Weertman    &    Anton Star**

University of Washington  
Department of Psychology, Neural Systems & Behavior Program  
Friday Harbor Laboratories

February 2026

*An individual-based simulation modeling the coupled dynamics of disease transmission, host evolutionary response, pathogen adaptation, and conservation intervention for the critically endangered sunflower sea star (*Pycnopodia helianthoides*) across 896 coastal sites spanning the northeast Pacific.*

# Contents

---

<b>1</b>	<b>Executive Summary</b>	<b>3</b>
<b>2</b>	<b>The Problem: Sea Star Wasting Disease</b>	<b>5</b>
2.1	A Catastrophic Decline . . . . .	5
2.2	The Causal Agent . . . . .	5
2.3	The Conservation Challenge . . . . .	5
<b>3</b>	<b>Model Architecture</b>	<b>7</b>
3.1	Individual-Based Simulation . . . . .	7
3.2	Disease Dynamics . . . . .	7
3.3	Three-Trait Genetic Architecture . . . . .	8
3.4	Pathogen Co-Evolution . . . . .	9
3.5	Temperature Forcing . . . . .	9
3.6	Calibration Approach . . . . .	9
<b>4</b>	<b>Spatial Network</b>	<b>10</b>
4.1	Network Structure . . . . .	10
4.2	Connectivity . . . . .	10
4.3	Real Satellite SST . . . . .	10
<b>5</b>	<b>Key Results</b>	<b>12</b>
5.1	The R→S Immunity Fix: A Paradigm Shift . . . . .	12
5.2	Resistance Is ~1000× More Valuable Than Tolerance . . . . .	12
5.3	The 3°C SST Window . . . . .	13
<b>6</b>	<b>Sensitivity Analysis</b>	<b>14</b>
6.1	Morris Screening (Round 4) . . . . .	14
6.2	Sobol Variance Decomposition (Round 4) . . . . .	14
6.3	Massive Interaction Effects . . . . .	14
6.4	Morris vs. Sobol Discrepancies . . . . .	15
<b>7</b>	<b>Conservation Genetics Module</b>	<b>16</b>
7.1	Scope and Structure . . . . .	16
7.2	Key Theoretical Results . . . . .	16
7.3	Five Analysis Templates . . . . .	17
<b>8</b>	<b>Performance and Scale</b>	<b>18</b>
8.1	Simulation Scale . . . . .	18
8.2	Computational Infrastructure . . . . .	18
8.3	Performance Optimizations . . . . .	18
8.4	Sensitivity Analysis Compute Budget . . . . .	19
<b>9</b>	<b>Current Status and Next Steps</b>	<b>20</b>
9.1	Project Status (February 2026) . . . . .	20
9.2	Immediate Next Steps . . . . .	20
9.3	Planned Scenario Suite . . . . .	20

<b>10 Technical Infrastructure</b>	<b>21</b>
10.1 Software Stack . . . . .	21
10.2 Codebase Metrics . . . . .	21

# 1 Executive Summary

---

Sea Star Wasting Disease (SSWD) has caused one of the most dramatic marine wildlife declines in recorded history, devastating populations of the sunflower sea star (*Pycnopodia helianthoides*) by over 90% across the northeast Pacific since 2013. As a keystone predator of sea urchins, the sunflower star’s collapse has triggered cascading ecosystem effects, including the expansion of urchin barrens that destroy kelp forests—a critical carbon sink and biodiversity hotspot.

**SSWD-EvoEpi** is a first-of-its-kind computational model that couples disease epidemiology, host evolutionary genetics, pathogen co-evolution, and spatial ecology to predict population recovery trajectories and evaluate conservation interventions. The model simulates **4.48 million individual sea star agents** across **896 coastal monitoring sites** spanning from Alaska to Baja California, driven by real satellite sea surface temperature data (NOAA OISST v2.1, 2002–2025) and CMIP6 climate projections to 2050.

## What the Model Does

At its core, SSWD-EvoEpi tracks every individual sea star through its complete life cycle—birth, growth, reproduction, infection, potential recovery, and death—while simultaneously tracking the evolution of three heritable immune traits (resistance, tolerance, and recovery) encoded across 51 genetic loci. The pathogen (*Vibrio pectenocida*) co-evolves in response to host adaptation, creating an eco-evolutionary feedback loop that determines whether populations can self-rescue through natural selection or require active conservation intervention.

## Why It Matters

Captive breeding programs are now underway at Friday Harbor Laboratories and the Sunflower Star Lab, with the first experimental outplanting of 28 captive-bred juveniles in 2025. **But where should we release them? How many? Should we selectively breed for disease resistance?** These are precisely the questions SSWD-EvoEpi is designed to answer—using mechanistic simulation rather than intuition.

## Key Results to Date

- **Sensitivity analysis** (960 Morris + 25,088 Sobol runs) has identified the critical parameter space: disease transmission parameters ( $K_{\text{half}}$ ,  $a_{\text{exposure}}$ ,  $P_{\text{env,max}}$ ) dominate population crash severity, with ~70% of output variance driven by parameter interactions.
- **Resistance is ~1000× more valuable than tolerance** for individual fitness at current population means—a quantitative prediction with direct implications for breeding program design.
- **The R→S immunity fix** (echinoderms lack adaptive immunity) fundamentally changes the model: recovery no longer confers lasting protection, shifting evolutionary pressure from clearance ability to infection avoidance.
- A narrow **3°C SST window (8.4–11.6°C)** drives the entire north–south recovery gradient via the VBNC (viable but non-culturable) pathogen sigmoid, explaining why Alaska populations are recovering while southern California populations are not.

- The companion **conservation genetics module** (2,600 lines, 146 tests) provides a complete theoretical framework for screening, breeding, and reintroduction optimization.

## 2 The Problem: Sea Star Wasting Disease

---

### 2.1 A Catastrophic Decline

Beginning in 2013, Sea Star Wasting Disease swept through the northeast Pacific in one of the largest marine wildlife die-offs ever documented. The disease manifests as tissue lesions that rapidly progress to limb autotomy (arm loss), body wall perforation, and death—often within days of initial symptoms. While SSWD affected over 20 asteroid species, the sunflower sea star (*Pycnopodia helianthoides*) suffered the most severe and geographically extensive decline.

- **>90% population decline** across the species' range, from Alaska to Baja California
- **Local extirpations** throughout southern portions of the range (California, Oregon)
- **No significant recovery** in most regions after a decade (2013–2023)
- **IUCN Critically Endangered** listing—the first for a sea star species
- **Cascading ecological effects:** urchin population explosions, kelp forest collapse, loss of biodiversity

### 2.2 The Causal Agent

The etiology of SSWD was debated for years, with early work implicating a densovirus (Hewson et al. 2014). However, a landmark study by **Prentice et al. (2025)** fulfilled Koch's postulates for *Vibrio pectenicida* as the causative agent of SSWD in *Pycnopodia helianthoides*. This bacterium is an environmental pathogen whose activity is strongly temperature-dependent—entering a viable but non-culturable (VBNC) state below a critical sea surface temperature threshold. This temperature dependence creates the latitudinal gradient in disease severity that is central to the SSWD-EvoEpi model.

### 2.3 The Conservation Challenge

With natural recovery stalled in much of the range, active conservation intervention has begun:

- **Friday Harbor Laboratories** (University of Washington): Captive rearing and disease challenge experiments
- **Sunflower Star Lab:** Captive breeding program with first outplanting trials in 2025 (28 captive-bred juveniles)
- **NOAA Species in the Spotlight:** Federal recovery planning

However, these programs face fundamental strategic questions:

1. **Where to release?** Which sites offer the best survival probability given local disease pressure and temperature regimes?
2. **How many to release?** What population supplementation scale is needed to overcome Allee effects and stochastic extinction?
3. **Should we breed for resistance?** Can selective breeding for disease-resistant genotypes accelerate population recovery, and at what cost to genetic diversity?

4. **Will climate change help or hurt?** Rising SSTs may shift pathogen activity patterns—but in which direction?

**SSWD-EvoEpi was built to answer these questions** through mechanistic simulation grounded in the best available empirical data.

### 3 Model Architecture

SSWD-EvoEpi is a spatially explicit, individual-based model (IBM) that couples four interacting subsystems: disease epidemiology, host evolutionary genetics, pathogen co-evolution, and population demography. This section describes each component and how they interact.

#### 3.1 Individual-Based Simulation

The model tracks **4.48 million individual agents** across 896 coastal sites. Each agent is a complete biological entity with the following tracked attributes:

Attribute	Type	Description
Position ( $x, y$ )	float32	Within-habitat coordinates
Size ( $L$ )	float32	Body length (cm), von Bertalanffy growth
Age	int16	Days since settlement
Sex	int8	Male/female
Disease state	int8	S, E, I <sub>1</sub> , I <sub>2</sub> , R, D
Disease timer	float32	Days remaining in current state
Resistance ( $r_i$ )	float32	Immune exclusion score $\in [0, 1]$
Tolerance ( $t_i$ )	float32	Damage limitation score $\in [0, 1]$
Recovery ( $c_i$ )	float32	Pathogen clearance score $\in [0, 1]$
Genotype	int8[51 × 2]	Diploid genotype at 51 loci

Each site maintains a population of up to ~12,500 individuals (carrying capacity  $K = 5,000$  with headroom for demographic fluctuations). The simulation advances in **daily timesteps** over a configurable time horizon (typically 13–20 years for calibration, up to 100 years for long-term projections).

#### 3.2 Disease Dynamics

Disease progression follows a compartmental model with temperature-dependent rates:

$$\begin{aligned}
 S &\xrightarrow{\lambda_i} E \xrightarrow{\mu_{EI_1}(T)} I_1 \xrightarrow{\mu_{I_1I_2}(T)} I_2 \xrightarrow{\mu_{I_2D}(T)} D \\
 I_1 &\xrightarrow{\rho_{\text{rec}} \cdot c_i} S \quad I_2 \xrightarrow{\rho_{\text{rec}} \cdot c_i} S
 \end{aligned}$$

where:

- **S** = Susceptible, **E** = Exposed (latent), **I<sub>1</sub>** = Early infected (mild wasting), **I<sub>2</sub>** = Late infected (severe wasting), **D** = Dead
- Recovered individuals return to **S** (not permanent immunity)—correct for echinoderms, which lack adaptive immune memory
- The force of infection for individual  $i$  is:

$$\lambda_i = a \cdot \frac{P}{K_{\text{half}} + P} \cdot (1 - r_i) \cdot S_{\text{sal}} \cdot f_{\text{size}}(L_i) \quad (1)$$

where  $P$  is environmental pathogen concentration,  $r_i$  is the individual's resistance score,  $S_{\text{sal}}$  is a salinity modifier, and  $f_{\text{size}}$  scales exposure with body size.

The environmental pathogen reservoir integrates contributions from shedding by infected ( $\sigma_1, \sigma_2$ ) and dead ( $\sigma_D$ ) animals, with temperature-dependent decay. Crucially, *Vibrio pectenicida* enters a **viable but non-culturable (VBNC)** state below a critical SST threshold, creating a temperature-driven on/off switch for disease activity:

$$f_{\text{VBNC}}(T) = \frac{1}{1 + e^{-k_{\text{VBNC}}(T - T_{\text{VBNC}})}} \quad (2)$$

This sigmoid function, parameterized by  $T_{\text{VBNC}} \approx 10^\circ\text{C}$ , produces a  $\sim 3^\circ\text{C}$  transition window ( $8.4\text{--}11.6^\circ\text{C}$ ) that maps directly onto the observed latitudinal recovery gradient: cold Alaska waters suppress pathogen activity, while warm southern waters sustain year-round transmission.

### 3.3 Three-Trait Genetic Architecture

The model encodes three distinct immune strategies across 51 diploid loci (based on the  $\sim 51$  loci under selection identified by Schiebelhut et al. 2018, 2024):

Trait	Loci	Symbol	Mechanistic Role
<b>Resistance</b>	17 (loci 0–16)	$r_i$	Immune exclusion. Reduces probability of infection. Receptor polymorphisms, barrier defenses.
<b>Tolerance</b>	17 (loci 17–33)	$t_i$	Damage limitation. Reduces mortality while infected but does not reduce shedding— <i>tolerant hosts are silent spreaders</i> .
<b>Recovery</b>	17 (loci 34–50)	$c_i$	Pathogen clearance. Increases probability of clearing infection and returning to S.

Each trait score is computed as the weighted sum of allelic effects:

$$r_i = \sum_{\ell=0}^{16} e_\ell \cdot \frac{g_{\ell,0} + g_{\ell,1}}{2}, \quad t_i = \sum_{\ell=17}^{33} e_\ell \cdot \frac{g_{\ell,0} + g_{\ell,1}}{2}, \quad c_i = \sum_{\ell=34}^{50} e_\ell \cdot \frac{g_{\ell,0} + g_{\ell,1}}{2} \quad (3)$$

where  $g_{\ell,k} \in \{0,1\}$  are alleles and  $e_\ell$  are locus-specific effect sizes drawn from an exponential distribution. Effect sizes are normalized so that each trait ranges  $[0,1]$ .

#### Key design principles:

- **One trait, one job:** Resistance reduces infection probability only. Tolerance extends  $I_2$  survival (via timer scaling) only. Recovery increases clearance probability only. No trait affects another’s mechanism.
- **Epidemiological consequences differ:** Resistance reduces both individual risk *and* population pathogen pressure. Tolerance saves the individual but maintains population transmission. Recovery removes shedding hosts but only after a period of infection.
- **The 17/17/17 partition is configurable** and can be explored in sensitivity analysis (two free parameters:  $n_r, n_t; n_c = 51 - n_r - n_t$ ).

Inheritance follows Mendelian segregation with mutation at rate  $\mu = 10^{-8}$  per locus per generation. The genotype array for the full simulation is  $(4.48\text{M} \times 51 \times 2) \text{ int8} \approx 460 \text{ MB}$ .

### 3.4 Pathogen Co-Evolution

The pathogen (*Vibrio pectenicida*) evolves virulence in response to host adaptation, creating a co-evolutionary arms race. Each pathogen lineage carries a continuous virulence trait  $v$  that governs trade-offs between:

- **Transmission** ( $\alpha_{\text{shed}}$ ): Higher virulence increases shedding rate
- **Kill rate** ( $\alpha_{\text{kill}}$ ): Higher virulence increases host mortality
- **Progression** ( $\alpha_{\text{prog}}$ ): Higher virulence accelerates disease progression

Virulence mutates each transmission event ( $\sigma_v$ ), allowing the pathogen to adapt to host resistance. This prevents unrealistic scenarios where hosts evolve resistance and the pathogen remains static.

### 3.5 Temperature Forcing

The model uses **real satellite sea surface temperature data** (NOAA OISST v2.1) from 2002–2025, providing daily SST values at each of the 896 model sites. For forward projections beyond 2025, the model ingests **CMIP6 climate model output** under multiple emissions scenarios (SSP1-2.6 through SSP5-8.5), enabling exploration of how climate change interacts with disease-evolution dynamics through 2050.

Temperature affects every major process: disease transmission rates, pathogen VBNC dynamics, growth rates, spawning phenology, and larval survival.

### 3.6 Calibration Approach

The model follows a five-phase calibration pipeline:

Phase	Purpose	Method	Runs
1	Sensitivity analysis	Morris + Sobol	~26,000
2	Parameter estimation	ABC-SMC	10,000–50,000
3	Convergence validation	Scale testing	~350
4	Scale correction	Bias adjustment	~200
5	Production scenarios	Reintroduction design	~500

Calibration targets include the observed 80–99% population crash, 2–5 year timeline from onset to nadir, north–south mortality gradient, fjord protection effects, and allele frequency shifts ( $\Delta q = 0.08$ – $0.15$ ) at immune loci (Schiebelhut et al. 2018).

## 4 Spatial Network

### 4.1 Network Structure

The SSWD-EvoEpi spatial network comprises **896 coastal monitoring and model sites** spanning the entire northeast Pacific range of *Pycnopodia helianthoides*, organized into 18 biogeographic regions:

ID	Region	Sites	SST Range	Habitat Type
1	Southeast Alaska	~60	5–12°C	Fjord-dominated
2	Haida Gwaii	~30	6–13°C	Mixed
3	Central BC Coast	~55	6–14°C	Fjord/open
4	Strait of Georgia	~45	7–15°C	Semi-enclosed
5	West Vancouver Island	~50	7–14°C	Open coast
6	Puget Sound	~70	8–15°C	Semi-enclosed
7	San Juan Islands	~35	7–13°C	Archipelago
8	Outer WA Coast	~40	8–14°C	Open coast
9	Oregon Coast	~55	8–15°C	Open coast
10	Northern California	~50	9–16°C	Open coast
11–18	Central–Baja California	~406	10–22°C	Mixed

### 4.2 Connectivity

Nodes are connected by **larval dispersal**, computed using an overwater distance matrix derived from a Dijkstra shortest-path algorithm on a high-resolution coastline graph. This approach:

- Ensures dispersal routes follow the coastline (no overland shortcuts)
- Captures the natural connectivity structure of the NE Pacific shelf
- Parameterizes dispersal probability as an exponential decay:  $C_{ij} = e^{-d_{ij}/D_L}$  where  $D_L$  is the characteristic larval dispersal distance
- Includes self-retention parameters that differ between fjord ( $\alpha_{\text{fjord}} \sim 0.8$ ) and open-coast ( $\alpha_{\text{open}} \sim 0.3$ ) sites

The pathogen also disperses between sites via a separate dispersal kernel (matrix **D**), representing waterborne pathogen transport with a maximum range of ~50 km.

### 4.3 Real Satellite SST

Each of the 896 sites is assigned a daily SST time series extracted from NOAA OISST v2.1 (0.25° resolution). This captures:

- Seasonal cycles (critical for spawning phenology and VBNC dynamics)
- Interannual variability (El Niño/La Niña, the 2013–2015 marine heatwave)
- The latitudinal temperature gradient (~5–22°C range across the network)
- Local anomalies (upwelling zones, fjord thermal buffering)

The 3°C VBNC transition window maps onto specific latitudinal bands in the network, creating the critical north–south divide: sites north of  $\sim 52^\circ\text{N}$  spend sufficient time below  $T_{\text{VBNC}}$  to suppress pathogen activity seasonally, while southern sites experience near-continuous disease pressure.

## 5 Key Results

### 5.1 The R→S Immunity Fix: A Paradigm Shift

One of the most consequential model refinements was correcting the post-recovery immune state. Echinoderms lack adaptive immunity—there is no immunological memory, no antibodies, no lasting protection after clearing an infection. The model was updated so that recovered individuals return to the **Susceptible** state (R→S) rather than retaining permanent immunity.

This single change fundamentally altered the model’s evolutionary dynamics:

Metric	Permanent Immunity	R→S (Correct)
Overall population crash	98.5%	99.7%
Final surviving population (5-node test)	365	122
Total disease deaths	41,968	36,157
Total recovery events	365	276
Nodes with local extinction	0/5	2/5

**Critical finding:** Under the corrected R→S model, the recovery trait *no longer evolves upward*. The maximum surviving-node change in recovery ability dropped from  $\Delta c = +0.154$  (permanent immunity) to  $\Delta c = +0.030$  (R→S)—a  $5\times$  reduction. Instead, **resistance replaces recovery as the dominant adaptive response**:

- Sitka (Alaska):  $\Delta r = +0.060$  under R→S vs.  $+0.011$  under permanent immunity ( **$5.5\times$  stronger**)
- Biological mechanism: When recovery doesn’t confer lasting protection, avoiding infection entirely (resistance) becomes far more valuable than clearing infection (recovery)

**Conservation implication:** Selective breeding programs should prioritize resistance traits over recovery ability.

### 5.2 Resistance Is $\sim 1000\times$ More Valuable Than Tolerance

The conservation genetics module derived marginal fitness derivatives at population-mean trait values, revealing a dramatic asymmetry:

Trait	Marginal Fitness Derivative	Relative Value
Resistance ( $r_i$ )	$\partial W / \partial r \approx 1000$	$1000\times$
Tolerance ( $t_i$ )	$\partial W / \partial t \approx 1$	$1\times$
Recovery ( $c_i$ )	$\partial W / \partial c \approx 50$	$50\times$

This asymmetry arises because resistance operates multiplicatively on the force of infection—a small increase in resistance reduces *every* exposure event, compounding across the individual’s lifetime. Tolerance, by contrast, only helps after infection has already occurred and only extends survival during  $I_2$ , without reducing pathogen transmission.

### 5.3 The 3°C SST Window

The VBNC sigmoid function creates a sharp environmental threshold. Within a narrow temperature window of  $\sim 8.4\text{--}11.6^\circ\text{C}$ , *Vibrio pectenica* transitions between dormancy and full activity. This 3°C window maps directly onto the observed latitudinal recovery gradient:

- **Cold sites** ( $\bar{T} < 8.4^\circ\text{C}$ ): Pathogen suppressed year-round. Populations can recover via demographics alone.
- **Transition zone** ( $8.4\text{--}11.6^\circ\text{C}$ ): Seasonal disease cycles. Partial recovery possible, especially in fjords with thermal buffering.
- **Warm sites** ( $\bar{T} > 11.6^\circ\text{C}$ ): Year-round disease pressure. Recovery requires evolutionary resistance or population supplementation.

This prediction is testable: sites near the VBNC threshold should show the greatest interannual variability in disease outcomes, driven by year-to-year SST fluctuations.

## 6 Sensitivity Analysis

A comprehensive sensitivity analysis was performed in two phases: Morris elementary effects screening (960 runs) followed by Sobol variance-based decomposition (25,088 runs), both on an 11-node stepping-stone network representing the full latitudinal range.

### 6.1 Morris Screening (Round 4)

The Morris method evaluates 47 parameters across 23 output metrics using one-at-a-time perturbations along randomized trajectories.

**Universal nonlinearity:** All 47 parameters show  $\sigma/\mu^* > 1.0$ , meaning every parameter participates in interactions with other parameters. This is a deeply coupled system with no purely additive effects. Three parameters show extreme interaction ratios ( $\sigma/\mu^* > 2.5$ ): pathogen mutation rate ( $\sigma_{v,\text{mut}}$ ) and the number of tolerance loci ( $n_{\text{tolerance}}$ ).

### 6.2 Sobol Variance Decomposition (Round 4)

The Sobol analysis used  $N = 512$  base samples with the Saltelli sampling scheme, producing 25,088 model evaluations executed in 20.4 wall-clock hours on a 48-core Intel Xeon W-3365. Results reveal a steep importance hierarchy:

Rank	Parameter	Module	$S_1$	$S_T$	Interpretation
1	$K_{\text{half}}$	Disease	0.182	<b>0.456</b>	Dose-response threshold
2	$a_{\text{exposure}}$	Disease	0.112	<b>0.337</b>	Transmission nonlinearity
3	$P_{\text{env,max}}$	Disease	0.049	<b>0.251</b>	Reservoir ceiling
4	$\sigma_{2,\text{eff}}$	Disease	0.041	<b>0.232</b>	Late-stage shedding
5	$\sigma_D$	Disease	0.075	<b>0.141</b>	Dead-animal shedding
6	$T_{\text{VBNC}}$	Disease	-0.015	0.040	Pathogen dormancy threshold
7	$k_{\text{growth}}$	Population	-0.009	0.035	Body growth rate
8	peak_width	Spawning	0.018	0.035	Spawning window width
9	settler_surv.	Population	0.010	0.033	Recruitment success
10	target_mean_r	Genetics	0.008	0.021	Initial resistance level

### 6.3 Massive Interaction Effects

The most striking finding is the dominance of parameter interactions. For population crash:

$$\frac{\sum S_T}{\sum S_1} = \frac{1.775}{0.527} = 3.37 \quad (4)$$

This ratio indicates that **~70% of total output variance arises from parameter interactions** rather than main effects. The top four disease transmission parameters form a tightly coupled cluster where each parameter's effect depends on the values of the others:

- $K_{\text{half}}$  and  $a_{\text{exposure}}$  jointly define the dose-response curve shape
- $P_{\text{env,max}}$  sets the exposure ceiling, modulating the dose-response relevance

- $\sigma_{2,\text{eff}}$  creates a shedding  $\rightarrow$  exposure  $\rightarrow$  infection  $\rightarrow$  shedding feedback loop

**Implication:** These parameters cannot be calibrated independently—joint estimation via ABC-SMC is essential.

## 6.4 Morris vs. Sobol Discrepancies

Notable rank changes between Morris and Sobol reveal the importance of global variance decomposition:

- $\rho_{\text{rec}}$  (recovery rate): Morris #1  $\rightarrow$  Sobol #32. Morris captured strong local effects, but Sobol shows these are absorbed by interactions.
- $\sigma_D$  (dead-animal shedding): Morris #20  $\rightarrow$  Sobol #5. The nonlinear feedback from post-mortem shedding is invisible to OAT perturbations.
- **Lesson:** Morris is effective for identifying the top  $\sim 10$  parameters but unreliable for precise ranking. Sobol is essential for calibration prioritization.

## 7 Conservation Genetics Module

A dedicated conservation genetics module provides the theoretical and computational framework for translating model outputs into actionable breeding program recommendations.

### 7.1 Scope and Structure

The module comprises **2,632 lines of code** across five submodules, supported by a 34-page theory report and validated by **146 tests** (136 pass; 10 fail for documented approximation reasons):

Module	Lines	Function
<code>trait_math.py</code>	221	Quantitative genetics: allele frequencies, additive variance, selection response, multi-generation predictions
<code>screening.py</code>	551	Sample size calculations for trait screening, exceedance probabilities, multi-site allocation, complementary founder selection
<code>breeding.py</code>	806	Mendelian crossing, four selection strategies (truncation, OCS, complementary, within-family), breeding program simulation
<code>inbreeding.py</code>	453	Genomic inbreeding coefficients, relationship matrices (VanRaden Method 1), effective population size estimation, inbreeding projections
<code>viz.py</code>	586	Comprehensive visualization suite

### 7.2 Key Theoretical Results

**Binary phenotyping framework.** In the absence of continuous phenotypic measurements for disease resistance, the module derives optimal screening strategies based on binary survival outcomes (survived vs. died during disease challenge). The theory shows how family-based selection and strategic crossing of survival categories can extract genetic information from coarse phenotypes.

**Breeding strategy comparison.** Four breeding strategies are implemented and compared:

- **Truncation selection:** Highest genetic gain ( $\Delta r = +0.71$  in 5 generations) but fastest erosion of additive variance  $V_A$
- **Optimal Contribution Selection (OCS):** Balances gain against inbreeding accumulation using genomic relationship matrix constraints
- **Complementary selection:** Lower per-generation gain but preserves allelic diversity across loci
- **Within-family selection:** Exploits high fecundity of *Pycnopodia helianthoides* ( $\sim 10^4$  eggs per female) for effective selection without reducing population-level diversity

**Practical constraint:** The module emphasizes that  $h^2$  is unknown for disease-related traits in *Pycnopodia helianthoides*. All generation-to-target estimates are reported as lower bounds (assuming  $h^2 = 1$ ), with guidance for scaling by  $1/h^2$  when empirical heritability estimates become available.

### 7.3 Five Analysis Templates

The module includes a complete analysis pipeline:

1. **Current genetic state:** Characterize post-epidemic trait distributions at 11 monitoring sites
2. **Screening effort:** Determine sample sizes needed to find high-resistance founders
3. **Breeding optimization:** Compare strategies across founder counts and generation sweeps
4. **Reintroduction scenarios:** Evaluate outplanting strategies with captive-bred stock (in development)
5. **Integrated recommendations:** Synthesize analyses into actionable conservation guidance

All templates are parameterized via a central `params.yaml` configuration and produce warnings when running with default (pre-calibration) parameters.

## 8 Performance and Scale

### 8.1 Simulation Scale

SSWD-EvoEpi operates at a scale that is unusual for ecological IBMs:

Dimension	Value
Total individual agents	4,480,000
Spatial nodes	896
Genotype array dimensions	$4.48\text{M} \times 51 \times 2$
Agent memory footprint	$\sim 1.7$ GB
Genotype memory footprint	$\sim 460$ MB
SST time series	$896 \times 4,745$ days
Distance matrix	$896 \times 896$
Daily timesteps per run (13 yr)	4,745
Total node-day operations per run	4,251,520

### 8.2 Computational Infrastructure

The primary compute platform is an **Intel Xeon W-3365** workstation:

- 64 physical cores (128 hardware threads) at 2.70 GHz
- 503 GB RAM
- Ubuntu 22.04 on WSL2

**Parallelization strategy:** Calibration runs are embarrassingly parallel—each parameter set evaluation is independent. By setting `OMP_NUM_THREADS=1` to disable per-process BLAS threading, the system runs up to 64 simultaneous simulation instances, fully utilizing all physical cores.

### 8.3 Performance Optimizations

A comprehensive performance audit identified the movement kernel (24 correlated random walk substeps per day per agent) as the dominant computational cost ( $\sim 60\text{--}70\%$  of total runtime). Key optimizations implemented or planned:

Optimization	Speedup	Status
Numba JIT for movement kernel	$17.5\times$	Implemented
Reduce substeps ( $24 \rightarrow 6$ )	$\sim 2\text{--}3\times$	Validated
Sparse pathogen dispersal matrix	$\sim 1.02\times$	Implemented
Parallel node processing (16 workers)	$\sim 4\text{--}8\times$	Planned
Pre-computed environmental arrays	$\sim 1.03\times$	Implemented

## 8.4 Sensitivity Analysis Compute Budget

Analysis	Runs	Wall Time	Platform
Morris R4 screening	960	~8 hours	Xeon (48 cores)
Sobol R4 decomposition	25,088	20.4 hours	Xeon (48 cores)
R→S validation	12	~0.5 hours	Local
<b>Total SA compute</b>	<b>26,060</b>	<b>~29 hours</b>	—

## 9 Current Status and Next Steps

### 9.1 Project Status (February 2026)

Component	Status	Completeness
Core simulation engine	Complete	100%
Three-trait genetic architecture	Complete	100%
Pathogen co-evolution module	Complete	100%
896-node spatial network + SST data	Complete	100%
Morris sensitivity analysis (R4)	Complete	100%
Sobol sensitivity analysis (R4)	Complete	100%
Conservation genetics module	Complete	95%
ABC-SMC calibration	In progress	30%
CMIP6 climate projections	Integrated	100%
Production reintroduction scenarios	Pending	0%
Manuscript	Pending	10%

### 9.2 Immediate Next Steps

1. **Complete ABC-SMC calibration** (6 rounds  $\times$  3 seeds) using the 15 priority parameters identified by Sobol analysis
2. **Convergence validation:** Verify that calibrated parameters produce consistent dynamics from  $K = 5,000$  to  $K = 100,000$  per node
3. **Scale correction:** Adjust any N-dependent parameters
4. **Production reintroduction scenarios:** Evaluate outplanting strategies (location, number, genetic composition) under baseline and climate-change projections

### 9.3 Planned Scenario Suite

Scenario	Seeds	Question
A. No intervention	20	Reference trajectory to 2050
B. Status quo (100 yr)	20	Long-term extinction risk
C. Current outplanting (50/yr, 5 sites)	20	Does current scale matter?
D. Aggressive outplanting (500/yr, 20 sites)	20	What scale is needed?
E. Genetic rescue (high- $r$ broodstock)	20	Does selective breeding help?
F. Assisted gene flow	20	Does connectivity management help?
G. Climate warming (+2°C)	20	How does warming shift disease dynamics?

## 10 Technical Infrastructure

---

### 10.1 Software Stack

Component	Technology
Core language	Python 3.10
Numerical computation	NumPy, SciPy
JIT compilation	Numba ( $17.5\times$ speedup on movement kernel)
Sensitivity analysis	SALib (Morris, Sobol)
Calibration	pyABC (ABC-SMC, planned)
SST data processing	xarray, netCDF4
Visualization	Matplotlib, Plotly
Version control	Git

### 10.2 Codebase Metrics

Component	Lines of Code
Core simulation model	$\sim 4,500$
Conservation genetics module	$\sim 2,600$
Analysis templates & scripts	$\sim 2,300$
Test suite	$\sim 3,500$
<b>Total</b>	<b><math>\sim 12,900</math></b>

- **770+ automated tests** covering disease dynamics, genetics, reproduction, spatial dispersal, conservation genetics, and integration
- **146 conservation genetics validation tests** with detailed reports linking each test to the corresponding equation in the theory document
- Full **Git version control** with structured commit history
- **CI-ready** test infrastructure

---

*For questions, collaboration inquiries, or access to the model code, please contact Willem Weertman at the University of Washington.*