**Parameters** of the model are in bold. *Variables* (vectors/matrices) are in italics.

*Step 1. Seed the model at the recruitment (pre-breeder) stage, just prior to competition for limited breeding sites.*

* Simulate **N** individuals with unique identifiers *1:N*, a fraction **F** of which are non-local intruders, the rest being local. So we initially have **F**­\***N** non-local individuals **and (1-F)\*N** local individuals.
* Next we set up 3 different genotype matrices, where rows hold individuals and columns hold loci. Possible allelic values for each cell are 0 and 1.
  + The first genotype matrix corresponds to the loci that uniquely affect *ZSOFT* (*Genotypes-\_soft\_unique*)
  + The second genotype matrix corresponds to loci that uniquely affect *ZHARD* (*Genotypes\_hard\_unique*),
  + The third genotype matrix corresponds to loci that are shared between both traits (i.e., affect both; *Genotypes\_both*).
* We then define a parameter for the desired total number of loci affecting each quantitative trait, **num\_loci\_total** (assumed to be the same for *ZSOFT* and *ZHARD*). We then define a parameter for the number of shared loci, **num\_loci\_shared**, that affect both traits (overlap between them). This allows us to calculate the derived parameter, **num\_loci\_unique = num\_loci\_total - num\_loci\_shared**, i.e., the number of loci uniquely affecting each trait.
  + So if we, for example, set **num\_loci\_total** = 30 and **num\_loci\_shared** =15, we then have 15 loci that uniquely affect each trait, and 15 that overlap between them (leading to a genetic correlation).
  + If we instead wanted completely uncorrelated traits, we simply set **num\_loci\_total** to 0.
* The number of rows for each of the above genotype matrices equals **N** (initial number of individuals).
  + The number of columns equals **2\*num\_loci\_unique** for the *Genotypes\_soft\_unique* and *Genotypes\_hard\_unique*matrices.
  + The number of columns equals **2\*num\_loci\_shared** for the *Genotypes\_both* matrix.
  + We multiply by 2 in each case because every individual carries two alleles per diploid locus. So the first two columns in each matrix correspond to Locus 1 (allele from inherited from “mum” and allele inherited from “dad”); columns 3 and 4 correspond to Locus 2…. Etc.
* We now populate these genotype matrices with 1s and 0s. We assume that every locus starts out at an expected allele frequency of 0.5, so for each cell of each matrix we simply toss a (fair) coin to see if gets a 1 or a 0.
  + Another way of doing this is draw the initial allele frequencies *pi* for each locus *i* from a beta distribution bounded between 0 and 1, of mean 0.5. Kardos et al. used such an approach in their eco-evo model, in order to allow for a range of initial allele frequencies across loci with whatever desired mean… not sure what we’d gain, in our case, from doing this, other than more parameters!)
* Next we column-bind the *Genotypes\_soft\_unique* and the *Genotypes\_both* matrices to generate an overall *Genotypes\_soft* matrix of genotypes affecting *ZSOFT.* Likewise, we bind *Genotypes\_hard\_unique* and *Genotypes\_both* to generate an overall *Genotypes\_hard* matrix of genotypes affecting *ZHARD*.
* On top of these, we create an additional genotype matrix to store alleles for the neutral trait (*Genotypes\_neutral*). The number of rows equals **N**, and the number of columns equals 2, as we assume only a single bi-allelic neutral locus.
  + This neutral locus is diagnostic with respect to local versus non-local, so local individuals are all assigned 0/0 genotypes and non-locals are all assigned 1/1 genotypes. The frequency of the 1 allele at this neutral marker in subsequent generations then allows us to keep track of introgression.
* The next step is to calculate the genotypic value (‘additive genetic merit’) of each (local or non-local) individual. This is done as follows, for *ZSOFT* and *ZHARD* separately:
  + Sum the number of 1 (trait-increasing) alleles across all loci for each individual.
  + For the soft-selected trait, multiply this sum by **a\_local\_soft** if the individual is local, and by **a\_nonlocal\_soft** if the individual is non-local.
  + For the hard-selected trait, multiply the corresponding genotypic values (sum of 1 alleles) by **a\_local\_hard** if the individual is local, and by **a\_nonlocal\_hard** if the individual is non-local.
  + These four parameters correspond to the per-locus genotypic effect sizes for each trait for each provenance. We assume additivity, i.e., no dominance or epistasis. We also assume that all loci have the same genotypic effect sizes.
    - So if say **a\_local\_soft** = 1, then at a single locus the genotypic effects for the 0/0 homozygote, 0/1 heterozygote and 1/1 homozygote, respectively, would be 0, 1 and 2. Summed across 30 loci, the overall multi-locus genotypic value of a local individual would then be somewhere between 0 and 60, with an expected (mean) value of 30 if both alleles are equally frequent.
    - If say **a\_nonlocal\_soft** = 0.5, then at a given locus the genotypic effects for the 0/0 homozygote, 0/1 heterozygote and 1/1 homozygote, respectively, would be 0, 0.5 and 1. Summed across 30 loci, the overall multi-locus genotypic value of a non-local individual would then be somewhere between 0 and 30, with an expected (mean) value of 15 if both alleles are equally frequent.
    - If **a\_nonlocal\_soft > a\_local\_soft**, then the mean genotypic value of non-locals would be higher than that of locals, hence locals would be at a competitive disadvantage.
  + Because we want full control over the initial phenotypic means of locals and non-locals for each quantitative trait (ZSOFT and ZHARD), we actually need to define four genotypic effect size parameters:
  + So we end up with an **N** x 2 matrix of genotypic values **G**, where the rows are individuals and columns are the two quantitative traits. The first **NF** rows are non-locals, whilst the second **N(1-F)** rows are locals.
  + The expected initial mean of these genotypic values for each group (provenance) equals n\*2\*a\*p, where n = number of loci affecting the trait, a = per-locus genotypic effect for that group, and p = initial allele frequency.
  + The expected initial variance of these genotypic values for each group (provenance) equals n\*2\*(a^2)\*p\*(1-p).
  + Thus if the genotypic effect (a) parameter for either trait is higher for locals than non-locals, then both the mean and the variance in genotypic values will be proportionately higher, but the evolvability (variance ÷ mean^2) will be the same. This is the way Kardos et al. did it.
* Now we draw environmental deviations for each individual for each quantitative trait and add them to the genotypic values (additive genetic merit) to arrive at the phenotypic value for that individual.
  + These environmental deviations are, by definition, centred on a mean of zero
  + Their variance equals the environmental variance**.**
  + We define an initial heritability **h2\_init** parameter. For simplicity, this is best kept at 0.5 across model runs.
  + The phenotypic variance Vp is the sum of the genetic and environmental variances, i.e. Vp = Vg + Ve, and the heritability h2 = Vg/Vp. So Ve = Vp – Vg = Vg/h2 – Vg. So we compute two initial environmental variances for each trait: one for the locals, and one for the non-locals:
    - **Env\_Var\_locals\_soft**= **Vg\_locals\_soft/h2\_init – Vg\_locals\_soft**
    - **Env\_Var\_nonlocals\_soft = Vg\_nonlocals\_soft/ h2\_init – Vg\_nonlocals\_soft**
    - **Env\_Var\_locals\_hard= Vg\_locals\_hard/ h2\_init – Vg\_locals\_hard**
    - **Env\_Var\_nonlocals\_hard= Vg\_nonlocals\_hard/h2\_init – Vg\_nonlocals\_hard**
  + The genetic variance (Vg\_) values in the above equations correspond to the expected initial (additive) genetic variances, given by n\*2\*(a^2)\*(0.5^2), where again n=number of loci, a = per-locus genotypic effect for that trait for that provenance, and p is the expected initial allele frequency per locus = 0.5. So these “expected initial genetic variances” are derived parameters, calculated via the inputted values for the total number of loci per trait (**num\_loci\_unique + num\_loci\_shared**) and the corresponding genotypic effects (**a\_local\_soft**, etc.). That in turn allows us to define the derived environmental variance parameters **Env\_Var\_locals\_soft**, etc.
  + So we can now draw our environmental deviations for the soft-selected trait for locals and non-locals from normal distributions of mean 0 and variances **Env\_Var\_locals\_soft** and **Env\_Var\_nonlocals\_soft**, respectively. There is no point in doing the same thing, at this step, for *ZHARD*, because the hard-selection filter comes after reproduction (and all we need at the reproduction step is the genotype matrix for *ZHARD*.)
  + And now we can define our phenotypic values for *ZSOFT* as the sum of the genotypic value and environmental deviation

*Step 2: Subject these initial individuals to soft selection*

* Rank the individuals from top to bottom based on their soft-selected trait, *ZSOFT*
* Define **K** as the number of breeding sites (habitat slots).
* If the number of pre-breeders > **K**, then select only the top **K** individuals from the right-tail of the *ZSOFT*distribution (based on their ranks). Kill off the other individuals.
* If the number of pre-breeders < **K**, then everyone gets a breeding site.
* In scenarios where soft selection is turned off (via the parameter **soft\_switch**, which is either “off” or “on”), just randomly select **K** breeders if the number of pre-breeders is > **K**.

*Step 3: Random mating and reproduction*

* We do not distinguish males and females, and we assume random hermaphroditic mating based on a classic Wright-Fisher model. This is pretty common in evolutionary models, and shouldn’t cause problems with reviewers, even though we notionally model a salmonid (not much to be gained by modelling separate sexes, if we don’t model any sexual dimorphism in our traits of interest). So each individual has an equal chance of becoming a parent, and each individual can produce more than 1 offspring (or no offspring). This guarantees an approximately Poisson distribution of offspring number per parent.
* The total number of offspring to be produced equals *Nb*\***k**, where *Nb* is the current number of breeders (*Nb*<=K) and **k** is fecundity (number of offspring per parent).
* Set up 3 new empty matrices to store the offspring genotypes for each quantitative trait and the neutral trait. So these have the same number of columns as the parental genotype matrices, and number of rows equal to the total number of offspring (*Nb*\***k**).
* To generate the first offspring, randomly draw its two parents from the vector of candidate parent IDs. Then, for each locus, flip a coin to choose which allele each parent passes on to that offspring. Remember, the first two columns of the parental genotype matrices for the quantitative traits correspond to the two alleles at locus 1, the next two columns correspond to the two alleles at locus 2, etc.
* Repeat this for all offspring, until the offspring genotype matrices have been populated with 1s and 0s (the inherited alleles).

*Step 4: Subject the offspring to hard selection*

* Now generate the phenotypes for Z­HARD for each individual by computing its genotypic value and adding an environmental deviation drawn from **Env\_Var\_locals\_hard** (as per above).
  + Note that all individuals now in this next generation are, by definition, locals, because they were born in the single wild environment considered.
  + Also note that we fix **Env\_Var\_locals\_hard** at its initial value as per Generation 1. This means that the realised heritability each generation might deviate from the initial 0.5 input value, because the realised genetic variance each generation will change owing to the level of introgression of non-local alleles, and erosion of genetic variance over time owing to selection and drift, whilst the (expected) environmental variance each generation will be constant.
* Now run the offspring through the hard-selection filter. So the expected survival *Wi* each individual is a function of its phenotype *ZHARD*, via a Gaussian fitness function defined by 3 parameters: *Wi* = **W­\_max**\*exp[-(*ZHARD* – **Theta**)/(2**\*Omega**2)]
  + **Theta** is the optimum value for ZHARD
  + **W\_max** is the maximum survival when *ZHARD* = the optimum.
    - To generate a stable population (prior to introgression), **W\_max** can be set to 1.5/**k**. So if **k** = 2 (each parent produces on average 2 offspring), then we set **W\_max** = 0.75. This ensures that an expected 1.5 offspring per parent survive the hard selection filter, i.e., recruits per spawner = 1.5, giving some reproductive excess.
  + **Omega** is the width of the fitness function, measured in units of *ZHARD*. Moderate selection here would correspond to an omega of about 5 times the phenotypic standard deviation (i.e., the expected initial standard deviation of *ZHARD* in this first generation of hard selection, which will vary across runs depending on the level of intrusion of non-locals and how different their mean *ZHARD* is relative to locals… so we will have to just choose a reasonable value for omega that applies across all these scenarios, i.e., **Omega** will be constant across runs but **Omega** ÷ standard deviation of *ZHARD* will not be constant).

*Step 5: Cycle back to Step 2, i.e., the survivors of hard selection are then the recruits (pre-breeders) that must compete for K limited breeding sites.*

* Back we go to the start. The survivors of hard selection now immediately graduate to pre-breeder status, and their genotypic matrices are now passed to the soft-selection filter.
* At this pre-breeder stage each generation, we calculate summary statistics of interest on a per-generation basis, as model output:
  + The phenotypic means and variances of *ZSOFT* and *ZHARD*
  + The genotypic variances for each. This then allows us to calculate the realised heritability for each.
  + The realised genetic correlation between *ZSOFT* and *ZHARD* (correlation between individual genotypic values for each)
  + In fact… it might be worth storing the entire matrices of individual genotypic values and phenotypes each generation, in order to explore how introgression affects the distributions (leads to initial bimodality, followed by skew??)
  + The allele frequency at the neutral (introgression marker) locus
  + The number of pre-breeders *N*.
  + The realised mean survival through the hard-selection filter (number of offspring just after divided by number of offspring just before).

Table 1. List of parameters. Does not include derived parameters, which can be computed based on these core parameters.

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter number** | **Symbol** | **Meaning** | **Suggested values to explore** |
| 1 | **N** | Initial number of pre-breeders | We aim for N=150 when the pop is composed of locals only (F=0), and intrusion then adds extra individuals.  So N = 150/(1-F).  E.g. if F=0.5, we have N=300 (150 locals and 150 non-locals), if F=0.25, we have N=200 (150 locals and 50 non-locals), etc. |
| 2 | **F** | Fraction of pre-breeders that are intruding non-locals (applied only in generation 1) | Explore a range of values from 0 to maybe 0.75. |
| 3 | **num\_loci\_total** | Total number of loci affecting each quantitative trait (*ZSOFT* and *ZHARD­*­­) | 30 |
| 4 | **num\_loci\_shared** | Number of loci affecting both traits | Set to 0 for scenarios where traits are uncorrelated, and set to somewhere between 0 and 30 for scenarios where traits are correlated. |
| 5 | **a\_local\_soft** | Genotypic effect per locus for *ZSOFT* of locals | Set to 1 always, as baseline |
| 6 | **a\_nonlocal\_soft** | Genotypic effect per locus for *ZSOFT* of non-locals | Explore a range of values <1 for scenarios where locals outcompete non-locals.  Explore a range of values >1 for scenarios where non-locals outcompete locals. |
| 7 | **a\_local\_hard** | Genotypic effect per locus for *ZHARD* of locals | Set to 1 always, as baseline |
| 8 | **a\_nonlocal\_hard** | Genotypic effect per locus for *ZHARD* of non-locals | Explore a range of values < 1. Non-locals are assumed to maladapted relative to locals, and it doesn’t matter in which direction. |
| 9 | **h2\_init** | Initial heritability of each trait | Fix at 0.5 in baseline case. In sensitivity analyses can explore a range of values between 0 and 1. |
| 10 | **K** | Number of breeding slots (carrying capacity) | Fix at 100 in baseline case, but vary across simulations in sensitivity analyses |
| 11 | **soft\_switch** | Turns soft selection on or off | ON or OFF |
| 12 | **k** | Fecundity (offspring per parent) | Fix at 2 |
| 13 | **W\_max** | Maximum height of hard-selection function (max offspring survival when *ZHARD* coincides with optimum | Fix at 0.75 |
| 14 | **Theta** | Optimum value for *ZHARD* | Fix at 30, which coincides with the expected phenotypic mean for *ZHARD* when the pop is entirely composed of locals only. |
| 15 | **Omega** | Width of fitness function, which determines strength of stabilising selection. | Fix at sqrt(150), which corresponds to 5 times the expected phenotypic standard deviation of *ZHARD* when the pop is entirely composed of locals only. |