

Full list of Specks configurable parameters			
Category	Parameter/Tag	Default	Description
Paths	output_folder_root	/home/SpecKS_output	Output destination. Write access required. SpeckKS will postpend a date and timestamp to the directory name, as in "SpeckKS_output_m04d26y2024_h12m21s55" so that subsequent runs are not overwritten.
Polyploid	name	[Auto/Allo]polyploid_N"	Name of polyploid. We suggest something descriptive.
Polyploid	SPC_time_MYA	0 MY	SPC time in MY. Time of subgenome divergence. For gradual speciation, this will be the mode of the gene tree divergence times.
Polyploid	WGD_time_MYA	0 MY	WGD time in MY. This will be the start time of ohnolog shedding, which will continue until the present time.
Polyploid	gene_div_time_distribution_parameters	impulse,1,1	Distribution of divergence times for gene trees at SPC time. For <b>autopolyploids</b> , use format: "impulse,1,1". For <b>allopolyploids</b> , use the format "expon,0,K", when K is the exponential decay constant. (We suggest $Ne \cdot Gt$ .). Lognormal distributions are also supported, with the format "lognorm,shape_parameter,xscale" (see <code>scipy.stats</code> , "lognorm" and "expon" for more details). For polyploids whose gene tree divergence might be a mix of distributions (ie, <b>segmental allopolyploids</b> ), multiple distributions may be given, with the last parameter being the proportion of genes which belong in each distribution. Ie, <gene_div_time_distribution_parameters> impulse,1,1,0.5 </gene_div_time_distribution_parameters> <gene_div_time_distribution_parameters> expon,0,10,0.5 </gene_div_time_distribution_parameters>
Species Tree	full_sim_time	100 MY	The length of the time period to simulate. Note that since speciation is a gradual process, it may be necessary to start the

			simulation well in advance of the SPC time.
Gene Tree	mean_gene_birth_rate_GpMY	0.001359 genes per MY	Reference "Gene family evolution in green plants with emphasis on the origination and evolution of Arabidopsis thaliana genes" - Ya-Long Guo
Gene Tree	SSD_half_life_MY	4 MY	"The Evolutionary Demography of Duplicate Genes" Michael Lynch, John S Conery
Gene Tree	WGD_half_life_MY	31 MY	Reference "Gene family evolution in green plants with emphasis on the origination and evolution of Arabidopsis thaliana genes" - Ya-Long Guo And (Maere et al., 2005)
Gene Tree	num_gene_trees_per_species_tree	3000	If set too low, the final histogram will look too sparse. Higher numbers may be more realistic (the number of arabidopsis genes is ~26000, ref How many genes are there in plants (... and why are they there)? 2007) and result in smoother histograms, but have a longer run time. Since gene trees are simulated independently, the number does not affect the general histogram shape, merely the number of samples in it, so high numbers are not always necessary.
Sequence Evolution	num_replicates_per_gene_tree	1	Specks can automatically run replicates for a given simulation, randomizing appropriately.
Sequence Evolution	num_codons	1000	The number of codons in each gene to be simulated. All genes in the sim have the same length. 1000 was chosen in agreement with Tiley 2018, and corroborated by "Long read sequencing to reveal the full complexity of a plant transcriptome by targeting both standard and long workflows"
Sequence Evolution	Ks_per_Myr	0.01	Ks per million years. This number is lineage and gene family specific, and may need to change depending on user needs (Gaut et al., 1996, (Koch et al., 2000)), . The default of 0.01 was chosen in agreement with Tiley 2018, and in range with[ Blanc, G. and Wolfe,K.H. 2004. Widespread paleopolyploidy in model plant

			species inferred from age distributions of duplicated genes. Plant Cell, 16, 1667-1678.]
Sequence Evolution	per_site_evolutionary_distance	0.01268182	The per-site-evolutionary_distance is used to calculate the total tree length per gene tree input to evolver. The default setting was derived by Tiley 2018, for a Ks_per_Myr of 0.01 and the evolutionary GY94 model.
Sequence Evolution	evolver_random_seed	137	The random seed used by evolver. We have exposed this so that the user can set this (or randomize it) directly as desired. The default value was simply chosen at random.
Main	LogFileName	log.txt	The name of the log file. The option to override the log file name supports the use case where the user might have multiple simultaneously executing runs sharing an output folder, and wishes them each to have a unique log file, to prevent overwriting or deadlock.
Histogram	maxKs	5	Debugging parameter. These histograms are only meant to give a confirmation of the run success.
Main	StopAtStep	999	Debugging parameter. If you want the sim to stop after only running the N'th module, set to N.
Main	IncludeVisuals	FALSE	Debugging parameter. If set to TRUE, .png files are generated for each gene tree at each stage. This can take up a lot of space on disk.

### Example Specks xml input file

```
<metadata>
  <Paths>
    <output_folder_root>/usr/scratch2/userdata</output_folder_root>
  </Paths>
  <SpeciesTree>
    <polyploid>
      <name>MyAlloPolyploid</name>
      <SPC_time_MYA>20</SPC_time_MYA>
      <WGD_time_MYA>18</WGD_time_MYA>
      <gene_div_time_distribution_parameters>expon,0,1.6
    </gene_div_time_distribution_parameters>
    </polyploid>
    <polyploid>
      <name>MyAutoPolyploid</name>
      <SPC_time_MYA>10</SPC_time_MYA>
      <WGD_time_MYA>10</WGD_time_MYA>
      <gene_div_time_distribution_parameters>impulse,1,1
    </gene_div_time_distribution_parameters>
    </polyploid>
    <polyploid>
      <name>MySegmentalPolyploid</name>
      <SPC_time_MYA>40</SPC_time_MYA>
      <WGD_time_MYA>38</WGD_time_MYA>
      <gene_div_time_distribution_parameters>expon,0,5,0.3
    </gene_div_time_distribution_parameters>
      <gene_div_time_distribution_parameters>impulse,1,1,0.7
    </gene_div_time_distribution_parameters>
    </polyploid>
    <full_sim_time>50</full_sim_time>
  </SpeciesTree>
  <GeneTree>
    <mean_gene_birth_rate_GpMY>0.07</mean_gene_birth_rate_GpMY>
    <SSD_half_life_MY>4</SSD_half_life_MY>
    <WGD_half_life_MY>31</WGD_half_life_MY>
    <num_gene_trees_per_species_tree>26000</num_gene_trees_per_species_tree>
  </GeneTree>
  <SequenceEvolution>
    <num_replicates_per_gene_tree>3</num_replicates_per_gene_tree>
    <num_codons>1000</num_codons>
    <Ks_per_Myr>0.01</Ks_per_Myr>
    <per_site_evolutionary_distance>0.01268182
  </per_site_evolutionary_distance>
  </SequenceEvolution>
  <Histogram>
    <maxKs>2</maxKs>
  </Histogram>
  <StopAtStep>999</StopAtStep>
  <LogFileName>SpecKS_log.txt</LogFileName>
  <IncludeVisuals>FALSE</IncludeVisuals>
</metadata>
```

