

# HostSwitch: An R Package to Simulate the Extent of Host-Switching by a Consumer

by Valeria Trivellone, Sabrina B. L. Araujo and Bernd Panassiti

**Abstract** In biology a general definition for host switch is when an organism (consumer) uses a new host (which represents a resource). The host switch process by a consumer may happen through its pre-existing capability to use a sub-optimal resource. The **HostSwitch** R package provides functions to simulate the dynamics of host switching (extent and frequency) in the population of a consumer that interacts with current and potential hosts over the generations. The **HostSwitch** package is based on a Individual-Based mock-up model published in FORTRAN by Araujo et al. (2015). The package largely improve the previous mock-up model, by implementing numerous new functionalities such as comparison and evaluation of simulations with several customizable parameters to accommodate several types of biological consumer-host associations, an interactive visualization of the model, an in-depth description of the parameters in a biological context. Finally, we provided three real world scenarios taken from the literature selected from ecology, agriculture and parasitology. This package is intended to reach researchers in the broad field of biology interested in simulating the process of host switch of different types of symbiotic biological associations.

## 1 Introduction

In several branches of biology (such as for example ecology, evolution, parasitology) a general definition for host-switching (or host shift) is when a consumer uses a newly colonized host, which represents its resource. Different spatial and temporal outcomes may result from the new host-consumer association depending on whether or not the colonization is successful. Studies on the evolution of biotic associations relies on an increasing body of literature covering all prototypical examples of symbiotic relationship categories (Thompson, 2010). Symbiosis *sensu lato* is defined here as any interaction between two organisms of different species. The possible influence between interacting organisms has been placed in a continuum of association types defined by their role, direction, and extension; and it varies from mutual consumption to unidirectional exploitation (Dimijian, 2000a,b). In the interspecific associations, the interacting organisms may play the role of strict consumer (e.g., predator) or resource (e.g., prey). Evolution of the associations may include several concatenated events of speciation affecting one or both species and it is driven by four main processes: cospeciation, host switch, failure to speciate and "missing the boat" (Page, 2002). While the prevalent paradigm considers cospeciation to be the main process driving evolution of most biological associations, recent evidence showed that, given the opportunity, a consumer may use a sub-optimal resource (or host) by host-switching without the need for any genetic innovation. This may explain the rapid origin of novel associations (i.e. colonization of novel hosts at the ecological time scale) eventually followed by speciation (at the evolutionary time) as well the observed incongruences of the paired phylogenies (see Brooks et al. (2019) for a review). Computer simulation modeling provide a valid tool to understand ecological and evolutionary dynamic of interacting species. To theoretically support the importance of host switch events, an Individual-Based Model (IBM) has been proposed by Araujo et al. (2015) (mock-up model hereafter). The model simulates the extent of host-switching in a host-parasite association formalized as the probability of an individual to disperse and successfully colonize a novel host. Recently, Feronato et al. (2021) provided a further add-on to the model by exploring the significance and the interaction of three parameters thought to be of paramount importance for the acquisition of a new host by a parasite. By using simulated data, these two initial papers provided important insights on the dynamic of host-switches for a parasite species. The main results were that host switch on a new host does not require prior evolutionary novelty, pathogens may survive on sub-optimal hosts which results in increased chance of host-switching to hosts more distant, and some parameters facilitate host switching (e.g., mutation and reproductive rate).

Both models were coded and run in FORTRAN language. Although available in an executable version (Windows, Linux, and MacOS), the previous mock-up model still lacks a user-friendly interface, and the manipulation of parameters is broadly limited. The current access to the model is restricted to certain groups of users who know how to compile and code in FORTRAN.

We present here a user-friendly R package, called **HostSwitch**, which improves the earlier version published in FORTRAN (Araujo et al., 2015; Feronato et al., 2021) in three main ways:

- (1) by increasing the accessibility of the model to researchers that are not familiar with FORTRAN;
- (2) by including customizable parameters, some previously unreleased (e.g., *jump\_back*), that allow us to extend the mock-up model from strict pathology (i.e., simulating host switches by pathogen) to

other branches of biology such as ecology, microbiology, agriculture (covering a very broader spectrum of symbiotic *sensu lato* associations, see Implementation and Usage scenarios sections for further details) reaching a broader audience;

(3) by providing in-depth descriptions of the parameters in a biological context, and examples of possible uses with real world data (see Usage scenarios section).

To our knowledge, there are no R packages that simulate the events of host switch using the theoretical approach briefly presented above and widely discussed in previous papers (Araujo et al., 2015; Agosta et al., 2010; Brooks et al., 2019). However, it is worth to mention that there are one putatively similar packages that may be used to simulate host switches. Notably, the package **EpiILM** (Vineetha Warriyar et al., 2020) uses discrete-time individual-level models to simulate the dynamic of infection disease transmission. The **EpiILM** package differs from the one present here for a fundamental point: **HostSwitch** simulate the host switch using the point of view of the consumer, rather than the host, and formalize the probability as random encounters of new hosts different to one other. For further details of the model formalize in **EpiILM** are presented in Deardon et al. (2010).

In Section "The model", we describe the mathematical framework for simulating the dispersion and the survival of individuals of a population on a novel resource which describes the event of host-switching. In Section "Implementation", we describe the implementation of the mock-up model in the **HostSwitch** package, including a description of the arguments of the main functions. In Section "Usage scenarios" we used empirical data gathered from the literature to simulate and compare different scenarios of host switch. We also present possible hypotheses and research questions that can be explored using the simulation approach.

## 2 The model

The simulation model of **HostSwitch** aims to measure the dynamics of host switching (extent and frequency) in the population of an organism (hereafter Consumer) that interacts with current and potential hosts (hereafter Resource) over generations. A successful host switch implies that a Consumer may colonize a new Resource, which in turn imposes selection pressure that impacts the Consumers' survival. The host-switching relies on a mechanism of ecological readjustment or ecological fitting, i.e. the capability of the Consumer to use a similar Resource even if sub-optimal (Janzen, 1985; Agosta and Klemens, 2008). The fundamental aspect of the **HostSwitch** simulation model is to track, summarize and compare the dispersion and successful host switch events in a new Resource by the populations of the Consumer. Although the model and the basic parameters have been previously described in Araujo et al. (2015) and Feronato et al. (2021), we provide here a revised description of the modeling dynamic which accommodates all symbiotic (*sensu lato*) associations.

The Resource is characterized by a real number,  $p_{Opt}$ , randomly selected from a uniform distribution ranging from  $p_{Res\_min}$  to  $p_{Res\_max}$ . This number represents the optimum phenotype imposed on Consumers by the Resource. Besides, the Resource imposes a carrying capacity  $K$  on the Consumer population. Individuals of the Consumer have a phenotype which can evolve over generations due to the emergence of novelties (hereafter mutations). Each individual consumer  $i$  is characterized by one phenotype  $p_i$ . The simulation starts (generation  $n = 0$ ) with all  $M$  consumer individuals having the same phenotype value ( $p_i = p_{Ind}$ ) which is equal to the average value in the resource range ( $(p_{Res\_min} + p_{Res\_max})/2$ ). At each generation, a novel Resource is offered and all Consumers have a probability  $mig$  to migrate from the current to the novel Resource. The number of dispersing individuals is calculated by assigning to each individual a value that follows the random uniform distribution. All the individuals with a value lower than  $mig$  disperse to the new Resource. Then, the assigned value may be interpreted ecologically as intrinsic or extrinsic characteristics involved in the dispersal event (e.g., morphological features, environmental constraints). The parameter  $mig$  defines a criterion for inclusion, with higher values allowing more individuals to disperse at each generation. The dispersion event has two possible outcomes: no migration (m1) and migration (m2).

- (m1) No individual disperses to the novel Resource and the model continues in the current Resource through the selection event. The Resource imposes on each individual a selection and the survival probability is calculated using the following equation:

$$P(p_i, p_{Opt}) = \exp\left[-\frac{(p_i - p_{Opt})^2}{2\sigma_{sel}^2}\right], \quad (1)$$

where  $p_i$  is the phenotype of  $i$ -th Consumer attempting to survive on the Resource;  $p_{Opt}$  is the optimum phenotype the Consumer should have to maximize survival success on the Resource;  $\sigma_{sel}$  is the standard deviation of a Gaussian distribution which is inversely related to the strength of selection. It may be interpreted as a proxy for the amplitude of the fitness space

of the Consumer, the higher the  $\sigma_{sel}$ , the lower the selection and the higher the probability of surviving. Ecologically this value may be related to the niche breadth for the Consumer.

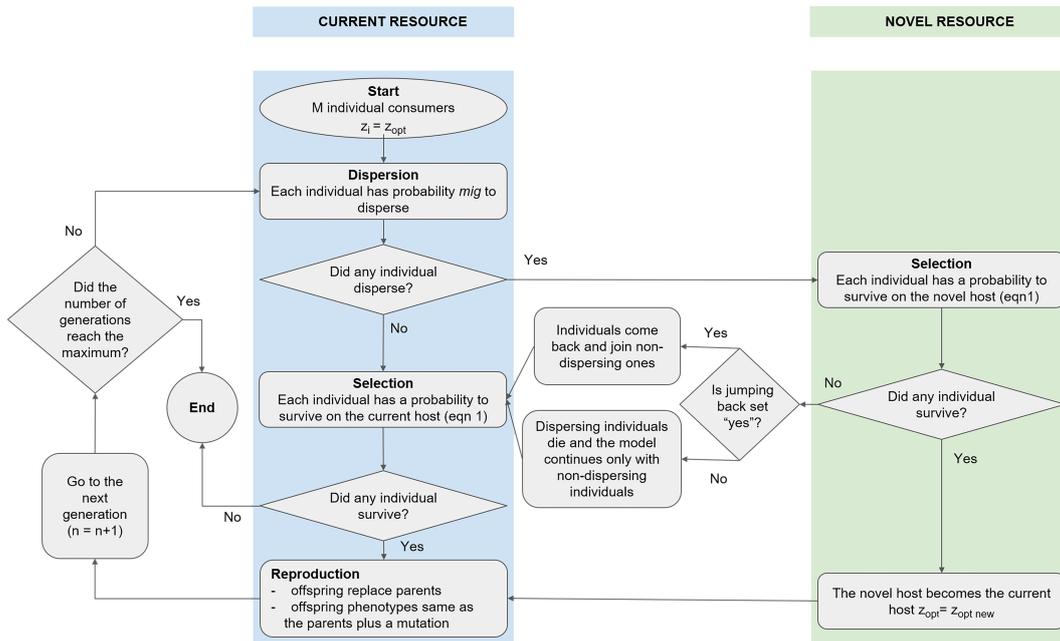
- (m2) Some individuals disperse and they go through a selection event on the novel Resource following eqn 1 (but with new  $p_{Opt}$  values, randomly defined). Here, the survival event has two possible outcomes: no survival (s1) and survival (s2).
  - (s1) None of the individuals survive and the model continues with the portion of individuals that did not migrate and are in the current (original) Resource, if dispersing individuals are not allowed to come back to the current Resource, or with individuals that did not migrate and those jumping back to the current Resource. The model imposes to these individuals a survival probability according to eqn 1.
  - (s2) Some individuals survive and the model continues only with those individuals.

All other individuals are ignored in subsequent steps. The novel Resource then becomes current Resource and the individuals will reproduce with a net reproduction rate  $b$ , limited to the carrying capacity  $K$ . The offspring's phenotype (inherit from the parents, plus variation  $\delta$ ) is assigned to each individual and calculated using the normal probability function:

$$P(\delta) = \exp\left[\frac{-\delta^2}{2\sigma_{mut}^2}\right], \tag{2}$$

where  $\delta$  is random phenotypic variation assigned to each individual;  $\sigma_{mut}$  is the standard deviation for mutation. The parameter  $\delta$  is randomly defined from a Gaussian distribution centered in zero and with a standard deviation  $\sigma_{mut}$ . Offspring phenotypes are equal to the arithmetic average of their parent's phenotypes plus  $\delta$  (i.e the extent of genetic novelty introduced with the reproduction). The descendants replace their parents and will populate generation  $n+1$  that will start over with another dispersion event to a novel Resource.

The overview of the main steps of the model are summarized in Figure 1.



**Figure 1:** Overview of the host-switching simulation model. At each generation every individual consumer has the chance to disperse to a novel Resource (on the right in green) and colonize it, or to jump back to the original Resource (on the left in blue). Selection (see eqn 1) is imposed by the current and the novel Resource. Individuals that survive the selection start a new generation after the reproduction event (see eqn 2). The framework is largely modified from Figure 1 after Araujo et al. (2015), more details in the text.  $z_i$  = phenotype of  $i$ -th Consumer;  $z_{opt}$  = phenotype that the Consumer should have to fit the novel Resource;  $z_{opt\ new}$  = phenotype that the Consumer retains on the current Resource after successful colonization of novel Resource;  $mig$  = a parameter defining the probability of dispersion of each Consumer.

### 3 Implementation

The **HostSwitch** R package is available for download from CRAN at <https://cran.r-project.org/web/packages/HostSwitch/index.html>. Latest versions of future developments of the R-package will be available on Github at <https://github.com/berndpanassiti/HostSwitch>.

This package provides an easy-to-use toolbox for users who want to simulate the host-switching of consumers and currently includes six functions (Table 1).

The package is installed and loaded by typing:

```
> install.packages("HostSwitch")
> library(HostSwitch)
```

**Table 1:** Description of the six functions in **HostSwitch** package

Function	Description
plotHostSwitch	Plots dispersal and colonization (host-switching events) of Consumers on a novel Resource offered at each generation given the values of parameters related to carrying capacity, fitness space, migration, reproduction, selection, and biological model.
shinyHostSwitch	Plots simulations of Consumer's host-switching on interactive web-based front-end using Shiny App.
simHostSwitch	Simulates the number of dispersion and successful host switch events by individuals of the Consumer until all individuals die.
summaryHostSwitch	Generates summary statistics for the object generated by simHostSwitch function.
survivalProbability	Includes the formula to calculate the survival probability used in the simHostSwitch function.
testHostSwitch	Tests the significance of the difference between two objects generated by simHostSwitch function.

The **HostSwitch** package is implemented around two main functions: `survivalProbability` and `simHostSwitch`. The former includes the formula of the survival probability (eqn 1), and the latter simulates host switches according to Figure 1 and encompasses dispersion, selection and reproduction events. `simHostSwitch` uses the `survivalProbability` function to calculate the probability of survival of the individuals on the current and new Resources.

The arguments of the `simHostSwitch` function are:

- *K*: carrying capacity, positive integer (min = 1, max = 1000), default value: 100.
- *b*: net reproduction rate; average number of offspring that a population of the Consumer produces at each generation, integer value (min = 0), default value: 10.
- *mig*: define the number of migrating individuals at each generation, numeric value (min = 0, max = 1), default value: 0.01. Larger values indicate higher probability to migrate.
- *sd*: standard deviation for mutation, numeric value (min = 0, max = 10), default value: 0.2. It corresponds to  $\sigma_{mut}$  in (eqn 2).
- *sigma*: standard deviation of the survival function, numeric value (min = 0, max = 10), default value: 1. It corresponds to  $\sigma_{sel}$  in (eqn 1).
- *pRes\_min*: smallest optimum phenotype value imposed by the Resource, numeric value (min = 1, max = *pRes\_max*), default value: 1.
- *pRes\_max*: highest optimum phenotype value imposed by the Resource, numeric value (min = *pRes\_min*, max = 100), default value: 10.
- *n\_generations*: number of generations, positive integer (min = 1, max = 50000), default value: 200.

- *jump\_back*: option for Consumers that do not survive on the novel Resource. If "yes" the consumer(s) is able to jump back to the current Resource and will be considered in the selective pressure and reproduction stage for the n+1 generation, if "no" (default) it cannot jump back and dies on the new host.
- *seed*: a single value useful for creating simulations or random objects that can be reproduced, positive integer (>0), default value: NULL.
- *n\_sim*: number of simulations, positive integer (min = 1, max = 50000), default value: 1.
- *nInitConsumer*: propagule size (or number of initial individuals) at the generation n = 0, default value: 20.
- *data, column*: instead of the arguments outlined above, the user may pass to the function a vector (*column* =) saved in a matrix or dataframe (*data* =) which specifies all or some of the arguments above, default value: NULL. The default value will be set if the argument is not provided. The arguments provided in dataframe will be overwritten if provided as an argument in the *simHostSwitch* function.

Using the *simHostSwitch* function, the R command below generates an object of class 'HostSwitch' which is a list of 18 elements, six simulated quantities and 12 used parameters (arguments of the function).

```
# generate a simulated 'HostSwitch' object using default values for the parameters
> sim1 <- simHostSwitch(n_sim= 3, seed = 2)
```

The simulated quantities are:

- *pRes\_sim*: a vector of the optimum phenotypes (one for each generation) that Consumers should have to be favored by the current Resource;
- *pRes\_new\_sim*: a vector of the optimum phenotypes (one for each generation) that Consumers should have to be favored by the novel Resource;
- *pInd\_sim*: list of vectors that includes the individual phenotype values of the Consumers in the population of each generation;
- *pInd\_jump\_sim*: vector of number of migrating individuals at each generation. The vector length is always equal to the 'n\_generation' parameter, if the simulation ends before the 'n\_generation' value then the vector will include a 'NA' by default;
- *pInd\_whichjump\_sim*: list of vectors that extracts the individual phenotype values of the Consumers who disperse in a novel Resource in each population and generation;
- *pInd\_whichsurv\_sim*: list of vectors that extracts the individual phenotype values of the Consumers who successfully colonize a novel Resource in each population and generation;

The function *summaryHostSwitch* creates a summary of basic statistics for phenotypes, dispersion and host switch events. The argument *warmup* defines the number of generations to be excluded from summary statistics. If *warmup* = 1 the generation at time 0 is excluded from summary, if *warmup* = 2 the generations at times 0 and 1 are excluded and so on. If *warmup* = NULL all generations are considered for summary statistics. Possible values are NULL or positive integer (min=1, max=50), default value = 1.

The example below provides the averages of the quantities of interest calculated for three simulations (*n\_sim* = 3)

```
# generate a summary for the simulated 'HostSwitch' object
> summaryHostSwitch(sim1)
An object of class summaryHostSwitch
Summary of HostSwitch simulations
```

General settings of individual based model:

```
K:100, b:10, mig:0.01, sd:0.2, sigma:1, pRes_min:1, pRes_max:10
n_generations:200, jump_back:no, seed:2, n_sim:3, warmup:1, nInitConsumer: 20
```

Summary of phenotypes:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
pRes	3.48	3.94	4.41	4.33	4.75	5.10
pRes_new	5.21	5.33	5.45	5.49	5.64	5.83
pInd	3.68	4.17	4.65	4.64	5.11	5.57

Summary of host switches by consumers:

	Mean	Max
Total events of dispersion:	29.67000	43
Number of successful host switches:	12.33333	16

The `testHostSwitch` function measures if there is a significant difference between the means of three simulated quantities of two `HostSwitch` objects (`simulated_quantities1` and `simulated_quantities2`). The argument `parameter` selects the quantity to compare, i.e. "j" (total number of dispersing events), "s" (total number of successful host switch events), and "d" (distance between the `pRes_sim` and `pRes_new_sim` for the generations where a successful host switch occurs, hereafter phenotype distance). The argument `test` defines the inferential statistic to be used, "t" parametric (t-test) or "w" nonparametric (wilcoxon rank test). The `warmup` is defined as above, and the argument `plot` provides the box plot of the results based on the `ggplot` function. When comparing quantities, the user needs to take into account all the arguments provided to the function `simHostSwitch`, including the number of simulations (`n_sim`). A warning message will be produced if `n_sim=1` for both `HostSwitch` objects. The following code generates another `HostSwitch` object (`sim2`), with `mig = 0.9` and default values for the others arguments. We compare the quantities from the objects `sim1` and `sim2` using `testHostSwitch`:

```
> sim2 <- simHostSwitch(mig = 0.9, n_sim = 3, seed = 3)

# compare the average values of the simulated quantities ("j", "s", "d") saved in the
# objects sim1 and sim2
> testHostSwitch(simulated_quantities1 = sim1, simulated_quantities2 = sim2,
  parameter = "j", test = "t", warmup = NULL, plot = FALSE)
An object of class testHostSwitch
Test result comparing 2 HostSwitch simulations using Welch Two Sample t-test
t: 2.74044, df:2.278635, p.value:0.09663877

> testHostSwitch(simulated_quantities1 = sim1, simulated_quantities2 = sim2,
  parameter = "s", test = "t", warmup = NULL, plot = FALSE)
An object of class testHostSwitch
Test result comparing 2 HostSwitch simulations using Welch Two Sample t-test
t: 2.505218, df:3.545114, p.value:0.0743971

> testHostSwitch(simulated_quantities1 = sim1, simulated_quantities2 = sim2,
  parameter = "d", test = "t", warmup = NULL, plot = FALSE)
An object of class testHostSwitch
Test result comparing 2 HostSwitch simulations using Welch Two Sample t-test
t: 1.271303, df:12.32786, p.value:0.2270841
```

The function returns the results of the test and the plot (if `plot = TRUE`). In the example above there is no significant difference between the two simulations, for all the simulated quantities ("j", "s", "d").

The `plotHostSwitch` and `shinyHostSwitch` functions both allow visualizing the results of a simulation, the first in the static plot window of the R console and the second one in a dynamic interface. The S3 method `plotHostSwitch` function graphically summarizes the simulated output. The following code generates the plot for the third simulation of the object `sim1`:

```
> plotHostSwitch(HostSwitch_simulated_quantities = sim1, n_sim = 3)
```

The function `shinyHostSwitch` relies on the package [shiny](#), and by running the code:

```
> shinyHostSwitch()
```

a popup window displays the plot for a simulation run with default values. The panel on the left includes slider bars for each argument of `simHostSwitch` function. The plot reacts dynamically to user input after pressing the Refresh simulation button.

## 4 Usage scenarios

In this section, we provide examples of how the functions described above can be applied to real data. We conducted a literature review with the aim to select studies providing data about the life cycle parameters of consumers interacting with their resources. The data were all gathered from

experimental settings and are related to three biotic interaction models representing distinct fields of research: wildlife ecology, agricultural pests, zoonotic pathogens. These data saved in the *parli* list object are included in the **HostSwitch** package. The list contains three matrices: `$Cephaloleia`, `$Cacopsylla`, and `$SarsMers`.

The complete code to reproduce results for the three biological scenarios presented is provided as an R Markdown file in the supplementary material for this article. Each scenario has two chunks of code: one to run the test and one to create plots. Using the Knit button in RStudio the results will be embedded beneath the code chunk in a .pdf document (see supplementary material for an example).

### ***Cephaloleia* - Zingiberales (wildlife ecology)**

Beetle species in the genus *Cephaloleia* (Coleoptera, Chrysomelidae, Cassidinae) are herbivorous insects feeding on monocots within the order Zingiberales and are known to be strictly co-evolved with their host plants during the last 35-60 Million years in the Neotropical region. This herbivore-plant association has been used earlier as a good model to test macroevolutionary hypotheses of host use and to investigate drivers of the evolution of the biological interactions (García-Robledo and Horvitz, 2011; Schmitt and Frank, 2013). Previous studies revealed host switch events in at least 7 species of *Cephaloleia* beetles in Costa Rica, and detailed information of life-cycle parameters have been reported (García-Robledo and Horvitz, 2011; Garcia-Robledo et al., 2010).

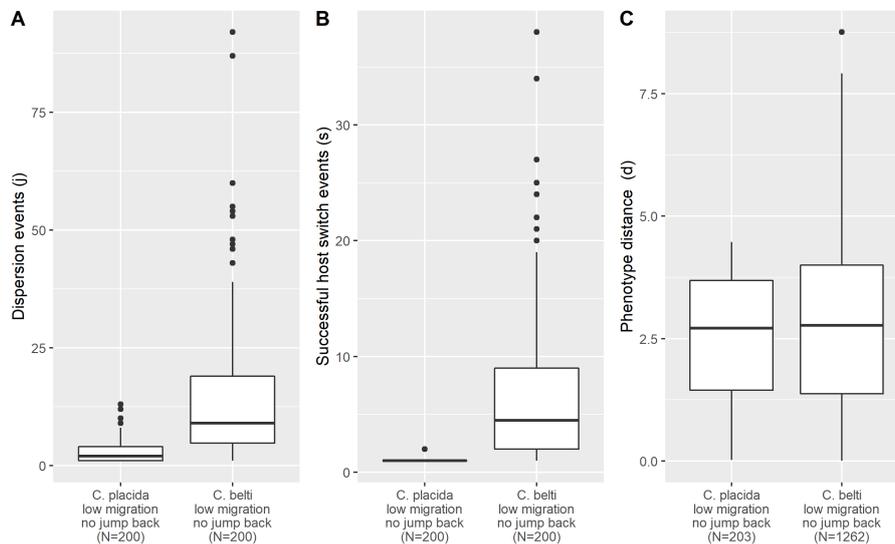
To inform our model, we used data on net reproduction rate for two species of *Cephaloleia* (the generalist *C. belti* and the specialist *C. placida*) provided by García-Robledo and Horvitz (2011). On their native host plants, *C. belti* showed the highest net reproduction rate ( $b = 4.6$ ) and *C. placida* the lowest ( $b = 1.4$ ) (see the lower value of the parameter for each species from Table 7 in García-Robledo and Horvitz (2011)). Suppose we are interested in simulating the host switch events of the two closely related taxa with different diet breadth, *C. belti* and *C. placida*. We may want to test the hypothesis that a higher net reproduction rate significantly increases the number of dispersion events and the successful colonization of new hosts (number of host switches). In this example, we simulated the dispersion and host switch events of the generalist *C. belti* (known to be associated with different families in Zingiberales in Costa Rica) and the specialist *C. placida* (hosted by *Renealmia alpinia*). For each species, the simulation was run for 200 generations,  $K=1000$  (carrying capacity not being a limiting factor), and 1000 runs. Because the other parameters, namely  $pRes_{min}$ ,  $pRes_{max}$ ,  $sd$  (standard deviation of mutation), and  $\sigma$  (standard deviation of survival), are not available in the literature for the two species the default values of the `simHostSwitch` function were used. The effect of the net reproduction rate on the simulated quantities (dispersion events  $j$ , successful host switches  $s$ , and distance between current and novel Resource  $d$ ) was tested, along with the assumption that  $b$  does not vary switching from the current to the novel resource across generations. We also tested the influence of the two parameters  $mig$  (migration probability) and  $jump\_back$  (possibility to jump back to the original host) with the assumption that they may rapidly vary under different ecological conditions. We arbitrarily set the variation for  $mig$  as low = 0.01 or high = 0.1, and for  $jump\_back$  as "no" or "yes". All possible combinations between the species were compared (Table 2).

The difference between the average values of  $j$ ,  $s$ , and  $d$  was evaluated with the paired t-test using the function `testHostSwitch`. The null hypothesis is that the pairwise difference between the two average values is equal. The code chunk "Scenario 1: Cephaloleia-Zingiberales (wildlife ecology)" (supplementary material) uses the data in `parli$Cephaloleia` object to generate a large list (`simResult`) with eight `simHostSwitch` objects, 4 for *C. belti* (hereafter Cb) and 4 for *C. placida* (Cp). The objects differ for the combinations of the parameters described above. The dataframe `testResult.Cephaloleia` saves the p-values for all the 16 combinations and all simulation objects defined by the user using the vector "simulations". All the objects are saved in the Global environment and a reshaped table with p-values of the t-test is visualized on the console and in Table 2. The code chunk "Plot for Cephaloleia" generates 3 plots for each of the tested quantities ( $j$ ,  $s$ , and  $d$ ); Figure 2 shows example box plots for the first comparison.

The results in Table 2 and Figure 2 show that the differences in the average number of dispersion and host switch events between *C. belti* and *C. placida* is significant, regardless of the values of probability of migration and the possibility to jump back to the old host. *Cephaloleia belti* shows higher values of  $j$  and  $s$ , this means that a four times higher net reproduction rate drives a higher probability of migration and successful host switches over time. As far as the distance between current and novel Resources is concerned, only seven out of 16 combinations showed a non-significant difference, all of them including *C. belti* with high probability of migration.

**Table 2:** Difference between the average values of three estimated quantities (total number of dispersing events  $j$ , total number of successful host switches  $s$ , phenotype distance  $d$ ) calculated from 200 simulations for each of the 16 independent combinations of two *Cephaloleia* species, (*C. belti* and *C. placida*). Fixed parameter from real data: net reproduction rate,  $b = 4.6$  *C. placida* and  $b = 1.4$  for *C. placida*. Arbitrary variation for two parameters:  $mig$ , probability of migration Low = 0.01 (default value) or High = 0.1;  $jump\_back$ , possibility to jump back to the original Resource No (default value) or Yes. Not significant p-values are shown in gray. For significant differences, the combination of the species with a higher value for the estimated quantities is shown in red.

<b>C. belti</b> (mig x jump_back)	<b>C. placida</b> (mig x jump_back)	<b>j</b>	<b>s</b>	<b>d</b>
Low x No	Low x No	<0.001	<0.001	<0.001
Low x No	High x No	<0.001	<0.001	<0.001
Low x No	Low x Yes	<0.001	<0.001	<0.001
Low x No	High x Yes	<0.001	<0.001	<0.001
High x No	Low x No	<0.001	<0.001	>0.05
High x No	High x No	<0.001	<0.001	<0.05
High x No	Low x Yes	<0.001	<0.001	>0.05
High x No	High x Yes	<0.001	<0.001	>0.05
Low x Yes	Low x No	<0.001	<0.001	<0.001
Low x Yes	High x No	<0.001	<0.001	<0.001
Low x Yes	Low x Yes	<0.001	<0.001	<0.001
Low x Yes	High x Yes	<0.001	<0.001	<0.001
High x Yes	Low x No	<0.001	<0.001	>0.05
High x Yes	High x No	<0.001	<0.001	>0.05
High x Yes	Low x Yes	<0.001	<0.001	>0.05
High x Yes	High x Yes	<0.001	<0.001	>0.05



**Figure 2:** Comparisons between the three simulated quantities (A- total number of dispersing events  $j$ , B- total number of successful host switches  $s$ , and C- phenotype distance  $d$ ) calculated for populations of *Cephaloleia placida* and *C. belti*. The populations are characterized by a low migration value and no possibility to jump back to the original Resource. The simulated quantities are calculated using the `simHostSwitch` function. The quantity  $d$  is defined as the difference between the phenotype of current and new Resource for generations where a successful host switch occurs. Each box plot depicts the distribution of the simulated quantities. For plot A and B,  $N$  represents the number of simulations (200). For plot C,  $N$  represents the total number of successful host switches calculated for 200 simulations.

### *Cacopsylla melanoneura* - Rosaceae (agricultural pests)

The hemipteran insect *Cacopsylla melanoneura* (Hemiptera, Psyllidae) is one of the known vectors of phytoplasmas, plant pathogenic bacteria in Class Mollicutes, and their association with host plants in the family of Rosaceae is known to be restricted to Central Europe (Janik et al., 2020). This group of pathogens is associated with severe diseases of orchards, including apple proliferation disease associated with '*Candidatus* (*Ca.*) Phytoplasma (*P.*) mali', pear decline ('*Ca.* *P. pyri*') and European stone fruit yellows ('*Ca.* *P. prunorum*'). The epidemiology of this vector-borne pathogen closely depends on the range of host plants that may serve as inoculum sources (reservoirs), and on the specific relationship between insect vectors and plants. Although *C. melanoneura* has been reported as a competent vector of '*Ca.* *P. mali*' in Italy (Tedeschi et al., 2002), the degree of preference for natural host plants may vary across countries affecting its role in phytoplasma spread, which is still largely unclear (Malagnini et al., 2010). Earlier investigations aimed to characterize aspects of its autoecology, such as life cycle, cryptic genetic variation, and habitat preference that may influence the associations with host plants (Malagnini et al., 2013). *Cacopsylla melanoneura* is known to be an oligophagous insect on hawthorn (*Crataegus* spp.), apple (*Malus* spp.), medlar (*Mespilus germanica*) and pear (*Pyrus communis*), and coniferous species are used as shelter plants during the winter (Jackson et al., 1990; Lal, 1934; Ossiannilsson, 1992). Moreover, the existence of *C. melanoneura* host races which show a morphological and genetic differentiation along with specific association with different host plants (apple or hawthorn) is likely in populations from different geographical regions (Lazarev, 1972, 1974; Lauterer, 1999; Malagnini et al., 2013). Available evidence suggests that sympatric speciation via host shift may be hypothesized. Even within the family Rosaceae only, the suite of host plants of *C. melanoneura* may vary locally through host switches with important consequences for the ecology of the insect vector and on the epidemiology of the phytoplasma disease in cultivated orchard trees.

We used the large amount of data available to inform our model and to compare two distinct populations of *C. melanoneura*: one adapted to apple (hereafter CmA) and the other to hawthorn (CmH). We used the experimental data provided by Malagnini et al. (2013) to calculate the net reproduction rate of the populations of *C. melanoneura* using eqn 2 of García-Robledo and Horvitz (2011). The calculated value  $b$  for CmA was about twice (2.196) that of CmH (1.022). To calculate the probability of migration  $mig$  we used the results of a choice test carried out using a dynamic olfactometer and provided in Figure 6 by Mayer et al. (2011). The values for  $mig$  for CmA were retrieved from the experiment with a population collected on apple tree that had a probability of 0.38 to migrate on apple. For CmH the probability was 0.49. The standard deviation of survival  $\sigma$  for CmA was retrieved by Malagnini et al. (2010) (see Results section) reporting a value of 3.68, and because no data was available for CmH we hypothesized a higher value 7 because hawthorn was reported as the ancestral host of *C. melanoneura* (Jackson et al., 1990).

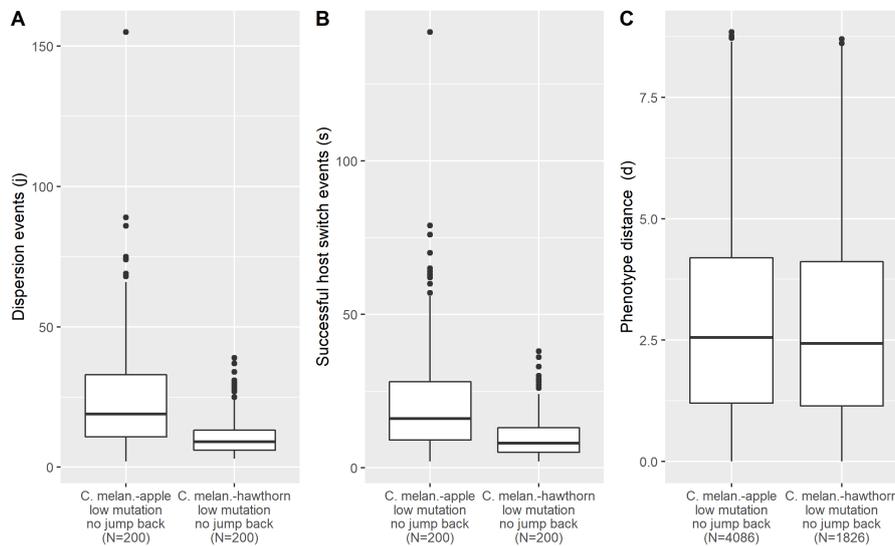
Suppose we are interested in simulating the host switch events of two populations of *C. melanoneura* to novel closely related hosts in family Rosaceae. We may want to test the hypothesis that twice a value of the standard deviation of survival  $\sigma$  for CmH is more important than a  $b$  two times higher for CmA, driving a significant increase in the number of dispersion events and more successful colonization of new hosts (number of host switches) for the population CmH than CmA. The parameter  $mig$  (migration) was also provided from real data but only slightly different between the two populations. Each simulation was run with 200 generations,  $K=1000$  (carrying capacity not being a limiting factor), and 1000 runs. The parameters  $pRes_{min}$  and  $pRes_{max}$  are not available in the literature and the default values were used. The effect of combined  $\sigma$  and  $b$  values on the simulated quantities (dispersion events  $j$ , successful host switches  $s$ , and distance between current and novel Resource  $d$ ) was tested, along with the assumption that both do not vary from the current to the novel Resource across generations. We arbitrarily set the variation for standard deviation of mutation  $sd$  as low = 0.2 or high = 9, and for the possibility to jump back to the original host  $jump\_back$  as "no" or "yes". All possible combinations between the populations were tested (Table 3).

We used the code chunk "Scenario 2: *Cacopsylla melanoneura*-Rosaceae (agricultural pests)" (supplementary file) to test the difference between the average values of simulated quantities ( $v_j$ ,  $s$ , and  $d$ ) for each combination of the populations of *C. melanoneura*. The null hypothesis is that the pairwise difference between the two average values is equal. The p-values of the t-test are reported in (Table 3. The code chunk "Plot for *Cacopsylla*" generates three plots for each of the tested quantities and in Figure 3 an example for the first combination is reported.

The results in Table 3 show that the difference in the average number of dispersion events is significant for all combinations except the one where we allowed the population adapted to hawthorn to come back to the original host and mutation is high for both populations. The comparisons of host switch events were significant except for three comparisons where mutation was set as high. When "j" and "s" are significantly different, the population of CmA shows higher values except for 5 combinations. These results suggest that a two times higher net reproduction rate drives a higher probability of successful host switch for CmA over time, masking the contribution of a two times

**Table 3:** Difference between the average values of three estimated quantities (total number of dispersing events  $j$ , total number of successful host switches  $s$ , phenotype distance  $d$ ) calculated from 200 simulations for each of the 16 independent combinations of two *Cacopsylla melanoneura* populations, *C. melanoneura* on Apple and *C. melanoneura* on Hawthorn. Fixed parameters from real data were: net reproduction rate,  $b = 2.196$ , probability of migration,  $mig = 0.37$ , and survival probability  $sigma = 3.68$  for *C. melanoneura* from apple and  $b = 1.022$ ;  $mig = 0.49$  and  $sigma = 7$  for *C. melanoneura* on hawthorn. Arbitrary variation for two parameters:  $sd$ , standard deviation of mutation Low = 0.2 (default value) or High = 9;  $jump\_back$ , possibility to jump back to the old Resource No (default value) or Yes. Not significant p-values are shown in gray. or significant differences, the combination of the species with a higher value for the estimated quantities is shown in red.

C. melanoneura Apple (sd x jump_back)	C. melanoneura Hawthorn (sd x jump_back)	j	s	d
Low x No	Low x No	<0.001	<0.001	>0.05
Low x No	High x No	<0.001	<0.001	>0.05
Low x No	Low x Yes	<0.001	<0.001	>0.05
Low x No	High x Yes	<0.001	<0.001	<0.01
High x No	Low x No	<0.01	<0.001	<0.01
High x No	High x No	<0.01	>0.05	>0.05
High x No	Low x Yes	<0.001	<0.001	<0.001
High x No	High x Yes	>0.05	<0.001	>0.05
Low x Yes	Low x No	<0.001	<0.001	>0.05
Low x Yes	High x No	<0.001	<0.001	>0.05
Low x Yes	Low x Yes	<0.001	<0.001	>0.05
Low x Yes	High x Yes	<0.001	<0.001	<0.01
High x Yes	Low x No	<0.001	<0.001	>0.05
High x Yes	High x No	<0.001	>0.05	>0.05
High x Yes	Low x Yes	<0.001	<0.001	<0.01
High x Yes	High x Yes	<0.001	>0.05	>0.05



**Figure 3:** Comparisons between the three simulated quantities (A- total number of dispersing events  $j$ , B- total number of successful host switches  $s$ , and C- phenotype distance  $d$ ) calculated for populations of *Cacopsylla melanoneura* (*C. melan.*) colonizing apple trees and for populations of *C. melanoneura* colonizing hawthorn. The populations are characterized by a low mutation value and no possibility to jump back to the original Resource. The simulated quantities are calculated using the `simHostSwitch` function. The quantity  $d$  is defined as the difference between the phenotype of current and new Resource for generations where a successful host switch occurs. Each box plot depicts the distribution of the simulated quantities. For plot A and B,  $N$  represents the number of simulations (200). For plot C,  $N$  represents the total number of successful host switches calculated for 200 simulations.

higher value of the standard deviation of survival. As far as the distance between current and novel Resources is concerned, 11 out of 16 combinations showed a non-significant difference, but no specific pattern has been revealed.

### ***Sarbecovirus* spp. and *Merbacovirus* sp. (SARS-MERS) - Mammals (zoonotic pathogens)**

The Severe Acute Respiratory Syndrome - Corona Virus (SARS-CoV) is a virus in the family *Coronaviridae* causing respiratory disease and was first identified in humans at the end of February 2003 in China. The Middle East Respiratory Syndrome - Corona Virus (MERS-CoV) (*Coronaviridae*, *Merbacovirus* sp.) was first isolated from humans in 2012 in Saudi Arabia. Lately, a SARS-like coronavirus (SARS-CoV-2, *Sarbecovirus* spp.) causing severe acute respiratory infections was first identified in December 2019 in Wuhan (China). The emergence of these viruses in humans appears to have been driven by stepping-stone switches (Parrish et al., 2008; Rodriguez-Morales et al., 2020), from bats to camels to humans, in the case of MERS-CoV and from bats to pangolins (among the others) to humans, in the case of SARS-CoV-2. As for many other human pathogens, to anticipate the emergence of future outbreaks is critical for appropriate measures to mitigate the consequences and costs associated with epidemic events.

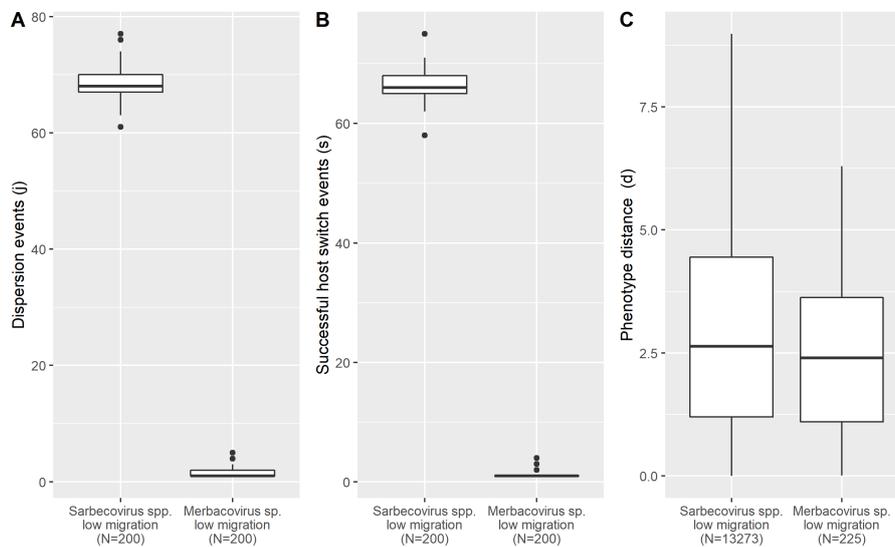
The usage scenario presented here offers an opportunity to apply our model to the biological association between viruses and their mammalian hosts. For this specific example we used the data provided by Kim et al. (2021) on viral dynamics of SARS-CoV-2, SARS-CoV, and MERS-CoV in humans. In particular we selected the two coronaviruses SARS-CoV-2 (hereafter Sars) and MERS-CoV (Mers), and two estimated population parameters: Maximum rate constant for viral replication and critical inhibition level (see Table 1 in Kim et al. (2021)), in our model corresponding to net reproduction rate ( $b$ , 4 for Sars and 1.46 for Mers) and standard deviation for survival ( $\sigma$ , 0.77 for Sars and 0.38 for Mers), respectively. The parameter  $\sigma$  was multiplied by 10 to be adjusted to the argument requirement of the model. Suppose we are interested in simulating host switch events of Sars and Mers among closely related mammalian hosts. We may test the hypothesis that higher net reproduction rate in combination with higher standard deviation of survival (i.e. lower selection on the Resource) significantly increases the number of dispersion events and successful colonization of new hosts (number of host switches). For each virus, the simulation was run with 200 generations,  $K=1000$  (carrying capacity not being a limiting factor), and 1000 runs. The parameter  $sd$  (standard deviation for mutation) was set at 0.002 for both viruses and was derived by the values of the substitution rate (per site per year) provided in Table 1 of van Dorp et al. (2020). The parameters  $pRes_{min}$ ,  $pRes_{max}$  and  $mig$  (migration) are not available in the literature and the default values were used. The combined effect of the net reproduction rate and survival probability on the simulated quantities was tested, along with the assumption that both parameters do not vary from the current to the novel Resource across generations. In this example, we also tested the influence of low (0.01), medium (0.5) and high (0.09) values of migration probability. The possibility to jump back to the original host ( $jump_{back}$ ) was set to "no" assuming that the virus is not able to come back from a new non-human mammal to human within the same generation (Table 4).

The code chunk "Scenario 3: Sarbecovirus sp. and Merbacovirus sp. (SARS-MERS)-Mammals (zoonotic pathogens)" (supplementary file) was used to test the difference between the average values of simulated quantities ( $v_j$ ,  $s$ , and  $d$ ) for each combination of the two viruses. The null hypothesis is that the pairwise difference between the two average values is equal. The p-values for all the 9 combinations are reported in (Table 4). The code chunk "Plot for SarsMers" generates three plots for each of the tested quantities and in Figure 4 an example for the first combination is reported.

The results in Table 4 and Figure 4 show that the differences in the average number of dispersion and host switch events between Sars and Mers are significant for all the simulations. The population of Sars shows higher values of the estimated quantities, except when Mers has high migration probability and Sars has low migration probability. This means that a combination of higher net reproduction rate and lower selection on the resource (both current and novel) drives to a higher probability of successful host switches over time. As far as the distance between current and novel resources is concerned, only 3 out of 9 combinations showed a non-significant difference, and for all these Mers has a high migration. In comparing the data from these coronaviruses, our purpose was merely academic and we did not mean to provide any evidence for the emergence of re-emergence of the present pandemic SARS-CoV-2 virus. Nevertheless, modeling support may certainly be of value when informed with detailed real data collected to address specific research questions.

**Table 4:** Difference between the average values of three estimated quantities (total number of dispersing events  $j$ , total number of successful host switches  $s$ , phenotype distance  $d$ ) calculated from 200 simulations for each of the 16 independent combinations of the two coronaviruses, Sars (*Sarbecovirus* sp.) and Mers (*Merbacovirus* sp.). Fixed parameters from real data were: net reproduction rate  $b = 4$ , standard deviation of survival  $\sigma = 7.7$ , and standard deviation of mutation  $sd = 0.002$  for Sars and  $b = 1.46$ ;  $\sigma = 0.38$  and  $sd = 0.002$  for Mers. Arbitrary variation for  $mig$ , probability of migration was set, Low = 0.01 (default value), Medium = 0.5 or High = 0.9. Not significant p-values are shown in gray. For significant differences, the combination of the species with a higher value for the estimated quantities is shown in red.

Sars (mig)	Mers (mig)	$j$	$s$	$d$
Low	Low	<0.001	<0.001	<0.001
Medium	Low	<0.001	<0.001	<0.001
High	Low	<0.001	<0.001	<0.001
Low	Medium	<0.001	<0.001	<0.001
Medium	Medium	<0.001	<0.001	<0.001
High	Medium	<0.001	<0.001	<0.001
Low	High	<0.001	<0.001	>0.005
Medium	High	<0.001	<0.001	>0.005
High	High	<0.001	<0.001	>0.005



**Figure 4:** Comparisons between the three simulated quantities (A- total number of dispersing events  $j$ , B- total number of successful host switches  $s$ , and C- phenotype distance  $d$ ) calculated for populations of *Sarbecovirus* spp. and *Merbacovirus* sp. The populations are characterized by a low probability of migration value. The simulated quantities are calculated using the `simHostSwitch` function. The quantity  $d$  is defined as the difference between the phenotype of current and new Resource for generations where a successful host switch occurs. Each box plot depicts the distribution of the simulated quantities. For plot A and B, N represents the number of simulations (200). For plot C, N represents the total number of successful host switches calculated for 200 simulations.

## 5 Conclusion and future avenues

This paper introduces the R package **HostSwitch** that provides functions to simulate, plot and test the number of dispersion and host switch events of a Consumer. It also measures phenotype distance between a current and novel Resource in the case of a successful host switch. **HostSwitch** is based on an earlier model published in FORTRAN (Araujo et al., 2015). The presented R-package offers several additional functions which largely improve the previous mock-up model, and accommodates users who want to simulate the ecological event of host-switching by using real parameters collected from different types of symbiotic biological associations. For example, the **HostSwitch** package can be used to test the hypothesis that host-switching happens more often than expected by chance.

The **HostSwitch** package was tested on real ecological Consumer-Resource interactions using data from the literature. We are aware that the examples provided above have several limitations, e.g., most of the arguments to inform the model were not available in the literature and the default values have been used. The reliability of the simulation on empirical data depends on the quality of the parameters used to inform the model and on their ecological significance. Users are therefore encouraged to set up specific experiments to collect parameters from a "host-switch perspective". The **HostSwitch** package will also continue to provide a theoretical foundation for understanding the specific processes that drive host switch events. Lastly, it represents a valuable user-friendly educational tool to facilitate deeper understanding of ecological and evolutionary dynamics. As a future avenue, we plan to extend the package to allow for the modeling of microbe-mediated tritrophic interactions such as occur in associations between pathogens, vectors and alternate hosts, e.g. in the phytoplasma pathosystem consisting of an insect vector, crop plants, and a pathogen that is able to manipulate its vector (Consumer) and host (Resource).

Albeit, the present package is intended for researchers in the broad field of biology who study consumer-host associations, we encourage users from other research areas to evaluate if the flowchart presented in Figure 1 may fit and be co-opted by other biological phenomena.

## 6 Acknowledgment

SBLA thanks Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) that provided partial funding. The authors thank Dr. Christopher H. Dietrich for constructive feedback on the final version of the paper.

## Bibliography

- S. J. Agosta and J. A. Klemens. Ecological fitting by phenotypically flexible genotypes: implications for species associations, community assembly and evolution. *Ecology Letters*, 11(11):1123–1134, 2008. URL <https://doi.org/10.1111/j.1461-0248.2008.01237.x>. [p180]
- S. J. Agosta, N. Janz, and D. R. Brooks. How specialists can be generalists: resolving the "parasite paradox" and implications for emerging infectious disease. *Zoologia (Curitiba)*, 27(2):151–162, 2010. ISSN 1984-4689. URL <https://doi.org/10.1590/S1984-46702010000200001>. [p180]
- S. B. L. Araujo, M. P. Braga, D. R. Brooks, S. J. Agosta, E. P. Hoberg, F. W. v. Hartenthal, and W. A. Boeger. Understanding host-switching by ecological fitting. *PLOS ONE*, 10(10):e0139225, 2015. URL <https://doi.org/10.1371/journal.pone.0139225>. [p179, 180, 181, 191]
- D. R. Brooks, E. P. Hoberg, and W. A. Boeger. *The Stockholm paradigm: climate change and emerging disease*. The University of Chicago Press, 2019. ISBN 978-0-226-63230-8 978-0-226-63244-5. URL <https://press.uchicago.edu/ucp/books/book/chicago/S/bo38871306.html>. [p179, 180]
- R. Deardon, S. P. Brooks, B. T. Grenfell, M. J. Keeling, M. J. Tildesley, N. J. Savill, D. J. Shaw, and M. E. J. Woolhouse. Inference for individual-level models of infectious diseases in large populations. *Statistica Sinica*, 20(1):239–261, 2010. ISSN 1017-0405. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4578172/>. [p180]
- G. G. Dimijian. Evolving together: the biology of symbiosis, part 2. *Proceedings (Baylor University Medical Center)*, 13(4):381–390, 2000a. ISSN 0899-8280. URL <https://doi.org/10.1080/08998280.2000.11927712>. [p179]
- G. G. Dimijian. Evolving together: the biology of symbiosis, part 1. *Proceedings (Baylor University Medical Center)*, 13(3):217–226, 2000b. ISSN 0899-8280. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1317043/>. [p179]

- S. G. Feronato, S. Araujo, and W. A. Boeger. 'Accidents waiting to happen'—insights from a simple model on the emergence of infectious agents in new hosts. *Transboundary and Emerging Diseases*, tbed.14146, 2021. ISSN 1865-1682. URL <http://onlinelibrary.wiley.com/doi/abs/10.1111/tbed.14146>. [p179, 180]
- C. García-Robledo and C. C. Horvitz. Experimental demography and the vital rates of generalist and specialist insect herbivores on native and novel host plants. *Journal of Animal Ecology*, 80(5):976–989, 2011. URL <https://doi.org/10.1111/j.1365-2656.2011.01843.x>. [p185, 187]
- C. Garcia-Robledo, C. C. Horvitz, C. L. Staines, et al. Larval morphology, development, and notes on the natural history of *Cephaloleia* "rolled-leaf" beetles (Coleoptera: Chrysomelidae: Cassidinae). *Zootaxa*, 2610:50–68, 2010. URL <https://doi.org/10.11646/zootaxa.2610.1.3>. [p185]
- C. Jackson, I. Hodkinson, P. Stanley, et al. Cold-hardiness in the hawthorn psyllid *Cacopsylla melanoneura* (Förster) (Homoptera: Psylloidea). *Entomologist*, 109(4):224–230, 1990. [p187]
- K. Janik, D. Barthel, T. Oppedisano, and G. Anfora. *Apple proliferation - A joint review*. Fondazione Edmund Mach/ Laimburg Research Center, 2020. ISBN 9788878430549. URL [http://www.laimburg.com/downloads/Apple\\_Proliferation.pdf](http://www.laimburg.com/downloads/Apple_Proliferation.pdf). [p187]
- D. H. Janzen. On ecological fitting. *Oikos*, 45:308–310, 1985. URL <https://doi.org/10.2307/3565565>. [p180]
- K. S. Kim, K. Ejima, S. Iwanami, Y. Fujita, H. Ohashi, Y. Koizumi, Y. Asai, S. Nakaoka, K. Watashi, K. Aihara, et al. A quantitative model used to compare within-host SARS-CoV-2, MERS-CoV, and SARS-CoV dynamics provides insights into the pathogenesis and treatment of SARS-CoV-2. *PLoS biology*, 19(3):e3001128, 2021. URL <https://doi.org/10.1371/journal.pbio.3001128>. [p189]
- K. Lal. The biology of Scottish Psyllidae. *Transactions of the Royal Entomological Society of London*, 82(2): 363–385, 1934. [p187]
- P. Lauterer. Results of the investigations on Hemiptera in Moravia, made by the Moravian museum (Psylloidea 2). *Acta Musei Moraviae, Scientiae Biologicae (Brno)*, 84:71–151, 1999. [p187]
- M. A. Lazarev. *Psylla melanoneura* frst. taurica forma nov. (Homoptera: Psyllidae). An apple tree pest in the Crimea. *Proceedings of the All-Union VI Lenin Academy of Agricultural Sciences, The State Nikita Botanical Gardens, Yalta, Ukraine*, 61:101–122, 1972. [p187]
- M. A. Lazarev. The 'Crimean Apple Sucker' *Psylla melanoneura* frst forma taurica, nov. (Homoptera, Psylloidea). *Nikita State Botanical Gardens, Ukraine*, pages 23–26, 1974. [p187]
- V. Malagnini, F. Pedrazzoli, V. Gualandri, F. Forno, R. Zasso, A. Pozzebon, and C. Ioriatti. A study of the effects of 'Candidatus Phytoplasma mali' on the psyllid *Cacopsylla melanoneura* (Homoptera: Psyllidae). *Journal of invertebrate pathology*, 103(1):65–67, 2010. URL <https://doi.org/10.1016/j.jip.2009.11.005>. [p187]
- V. Malagnini, F. Pedrazzoli, C. Papetti, C. Cainelli, R. Zasso, V. Gualandri, A. Pozzebon, and C. Ioriatti. Ecological and genetic differences between *Cacopsylla melanoneura* (Homoptera, Psyllidae) populations reveal species host plant preference. *PloS one*, 8(7):e69663, 2013. URL <https://doi.org/10.1371/journal.pone.0069663>. [p187]
- C. J. Mayer, A. Vilcinskis, and J. Gross. Chemically mediated multitrophic interactions in a plant–insect vector–phytoplasma system compared with a partially nonvector species. *Agricultural and Forest Entomology*, 13(1):25–35, 2011. URL <https://doi.org/10.1111/j.1461-9563.2010.00495.x>. [p187]
- F. Ossiannilsson. *The Psylloidea (Homoptera) of Fennoscandia and Denmark*, volume 26. Brill, 1992. [p187]
- R. D. M. Page. Introduction. In *Tangled Trees. Phylogeny, Cospeciation and Coevolution*, pages 1–22. The University of Chicago Press, 2002. URL <https://press.uchicago.edu/ucp/books/book/chicago/T/bo3634552.html>. [p179]
- C. R. Parrish, E. C. Holmes, D. M. Morens, E.-C. Park, D. S. Burke, C. H. Calisher, C. A. Laughlin, L. J. Saif, and P. Daszak. Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiology and Molecular Biology Reviews*, 72(3):457–470, 2008. URL <https://doi.org/10.1128/MMBR.00004-08>. [p189]
- A. J. Rodriguez-Morales, D. K. Bonilla-Aldana, G. J. Balbin-Ramon, A. A. Rabaan, R. Sah, A. Paniz-Mondolfi, P. Pagliano, and S. Esposito. History is repeating itself: Probable zoonotic spillover as the cause of the 2019 novel coronavirus epidemic. *Infez Med*, 28(1):3–5, 2020. [p189]

- M. Schmitt and M. Frank. Notes on the ecology of rolled-leaf hispines (Chrysomelidae, Cassidinae) at La Gamba (Costa Rica). *ZooKeys*, 332:55, 2013. URL <http://dx.doi.org/10.3897/zookeys.332.5215>. [p185]
- R. Tedeschi, D. Bosco, and A. Alma. Population dynamics of *Cacopsylla melanoneura* (Homoptera: Psyllidae), a vector of apple proliferation phytoplasma in northwestern Italy. *Journal of economic entomology*, 95(3):544–551, 2002. URL <https://doi.org/10.1603/0022-0493-95.3.544>. [p187]
- J. N. Thompson. Four central points about coevolution. *Evolution: Education and Outreach*, 3(1): 7–13, 2010. ISSN 1936-6434. doi: 10.1007/s12052-009-0200-x. URL <https://evolution-outreach.biomedcentral.com/articles/10.1007/s12052-009-0200-x>. Number: 1 Publisher: BioMed Central. [p179]
- L. van Dorp, M. Acman, D. Richard, L. P. Shaw, C. E. Ford, L. Ormond, C. J. Owen, J. Pang, C. C. Tan, F. A. Boshier, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics and Evolution*, 83:104351, 2020. URL <https://doi.org/10.1016/j.meegid.2020.104351>. [p189]
- K. V. Vineetha Warriyar, A. Waleed, and D. Rob. Individual-Level Modelling of Infectious Disease Data: EpiILM. *The R Journal*, 12(1):87–104, 2020. doi: 10.32614/RJ-2020-020. URL <https://doi.org/10.32614/RJ-2020-020>. [p180]

*Valeria Trivellone*  
*University of Illinois at Urbana-Champaign*  
*Prairie Research Institute*  
*1816 South Oak Street, Champaign, Illinois 61820*  
*United States of America*  
ORCID: <http://orcid.org/0000-0003-1415-4097>  
[valeria.trivellone@gmail.com](mailto:valeria.trivellone@gmail.com)

*Sabrina B. L. Araujo*  
*Universidade Federal do Paraná*  
*Departamento de Física*  
*81531-980 Curitiba, Paraná*  
*Brazil*  
*Laboratório de Ecologia e Evolução de Interações*  
*Biological Interactions Universidade Federal do Paraná*  
*P.O. Box 19073 PR 81531-980, Curitiba*  
*Brazil*  
ORCID: <https://orcid.org/0000-0002-8759-8310>  
[araujosbl@gmail.com](mailto:araujosbl@gmail.com)

*Bernd Panassiti*  
*Independent researcher*  
*Munich*  
*Germany*  
ORCID: <http://orcid.org/0000-0002-5899-4584>  
[bernd.panassiti@gmail.com](mailto:bernd.panassiti@gmail.com)