# Genomic patterns of divergence during adaptive speciation

Raphaël Scherrer[1]        Ulf Dieckmann[2]        G. Sander van Doorn[1,*]

1. Groningen Institute for Evolutionary Life Sciences, University of Groningen, the Netherlands;

2. International Institute for Applied Systems Analysis

* Corresponding author; e-mail: g.s.van.doorn@rug.nl

# Abstract

xxx

# Introduction

# Model

Building on previous work by Ripa *et al.*, we model a population of consumers that can feed on two alternative resources in two habitat patches connected by migration. An individual's ecological trait $x$ controls its feeding efficiencies on the two resources, $e_1(x)$ and $e_2(x)$, which are subject to a trade-off. Depending on the trade-off strength, the distribution of resources over the two habitats and the amount of gene flow between them, selection can induce the ecological trait to undergo adaptive diversification. However, in a sexually reproducing species, recombination tends to thwart the evolution of distinct resource specialists, unless a reproductive isolation barrier evolves to prevent hybridisation. The basis for reproductive isolation in our model is mate choice by females, which can evolve on a continuous scale extending from random mating to both positively and negatively assortative with respect to the ecological trait.

As we are interested in characterising the genomic signatures of different modes of speciation, as well the dependence of these patterns on the genomic architecture of speciation traits, we develop a detailed model of the genotype-phenotype map, enabling a systematic investigation of the consequences of gene-linkage and non-additive genetics. The output of the simulations provides high-resolution data, comparable to genome sequences, from which we extract information by means of methods taken from speciation genomics. A summary explanation of the different components of the model is provided in the sections below. Additional information can be found in the supplementary information.

### Divergent ecological selection

The two resources that occur in each habitat are modelled as abiotic resources; each enters the system at a constant rate ($I_i^{[j]}$ for resource $i$ in habitat $j$) and is washed out at rate $c_i^{-1} R_i^{[j]}$ (i.e., directly proportional to its abundance, $R_i^{[j]}$, where $c_i$ is the renewal time for resource $i$). In addi-

tion, resources are taken up by the consumers at rate $e_i(x_k)\,R_i^{[j]}$ for each consumer $k$ who feeds on the resource. Consumers do not utilise both resources simultaneously, but are assumed to switch adaptively between consuming one or the other, in order to maximise their instantaneous resource uptake rate. Accordingly, the resource dynamics are described by

$$\frac{d\,R_1^{[j]}}{dt} = I_1^{[j]} - R_1^{[j]}\left(c_1^{-1} + \sum_{k<\theta_j} e_1(x_k)\right) \tag{1}$$

$$\frac{d\,R_2^{[j]}}{dt} = I_2^{[j]} - R_2^{[j]}\left(c_2^{-1} + \sum_{k>\theta_j} e_2(x_k)\right) \tag{2}$$

Here, the consumers are taken to be ranked (from high to low) with respect to their relative feeding efficiency on resource 1 and $\theta_j$ is a number chosen in each habitat such that $e_1(x_k)R_1^{[j]} > e_2(x_k)R_2^{[j]}$ for all consumers with index $k < \theta_j$. Note that differences between habitats are created exclusively by asymmetries in the resource influx rates $I_i^{[j]}$ and not by potential habitat-specific resource renewal times $c_i$ or feeding efficiency functions $e_i$. Incorporating such differences is straightforward, but here we prefer to avoid additional complexity in the ecological model.

Under the assumption that the resource dynamics are fast relative to the population dynamics of the consumer, we calculate the feeding rate by each individual based on the resource equilibrium abundances $\hat{R}_i^{[j]}$. We then take this measure as a proxy for the fertility $f_k^{[j]}$ of individual $k$ in habitat $j$, i.e.,

$$f_k^{[j]} = \begin{cases} e_1(x_k)\,\hat{R}_1^{[j]} = b_1^{[j]}\,\dfrac{e_1(x_k)}{1+c_i\sum_{\kappa<\theta_j} e_1(x_\kappa)} & \text{if } k < \theta_j \\[2ex] e_2(x_k)\,\hat{R}_2^{[j]} = b_2^{[j]}\,\dfrac{e_2(x_k)}{1+c_i\sum_{\kappa>\theta_j} e_2(x_\kappa)} & \text{otherwise} \end{cases} \tag{3}$$

where $b_i^{[j]} = c_i\,I_i^{[j]}$ is the maximum per capita offspring production rate by consumers feeding on resource $i$ in habitat $j$. Death of consumer individuals occurs at a constant per capita rate $d$ (independent of feeding behaviour and habitat).

## Mate choice

The reproductive success of both males and females is determined by their fertility and their success in attracting and finding a mate. We assume that females exert mate choice. Depending on the value of her mating trait $y_k$, a female with index $k$ either prefers to mate with a male that has an ecological trait value similar to her own (positively assortative mating; $y_k > 0$), a male that has a different ecological trait value (negatively assortative mating; $y_k < 0$), or she selects a mate at random ($y_k = 0$).

A fertile female encounters potential mates one-by-one and can each time either reject or accept the current male as a mating partner. Previously rejected candidates cannot be revisited. Each male encountered by the female is sampled with replacement from her local habitat, according to a weighted-lottery with weights equal to the male fertility values $f_m^{[j]}$ (for all males $m$ in habitat $j$). After accepting a mate, the female produces a single offspring. If she can produce additional offspring, she continues to search for mates until her maximum reproductive success has been realised or the mating season ends. The total number of offspring that can be produced by female $k$ during the current reproductive season follows a Poisson distribution with mean equal to the female's fertility $f_k^{[j]}$. The length of the mating season is drawn from a geometric distribution; the number of potential mates that can be assessed by a female before the season ends is a random variable $N$, with $\Pr[N = n] = q\,(1 - q)^n$. Parameter $q$ quantifies the cost of mate choice and corresponds to the probability that a female, after rejecting a potential mate, will not be able to find another male before the season ends. In line with many other models of mate choice, we assume that $q \ll 1$, so that direct selection on female choosiness is weak.

We assume that females employ an optimal threshold rule to decide between accepting or rejecting a potential mate. Female $k$ accepts male $m$ as a mate only if the score $u_{k \times m}$ attributed to the male by the female exceeds the female's individual quality threshold $u_k^*$. We define the male scores as:

$$u_{k \times m} = S(x_k - x_m, y_k) = -a\,y_k \left( \frac{x_k - x_m}{\sigma} \right)^2 \tag{4}$$

5

Here, $a$ is a parameter that scales the strength of mate preferences, $y_k$ is the female mating trait and $\sigma$ is equal to the standard deviation of ecological traits (in the local population). Note that the male scores are subjective; they depend on the female mating preference as well as on the normalised difference between male and female ecological traits.

The female's threshold for mate acceptance, $u_k^*$, reflects a compromise between the costs and the (perceived) benefits of rejecting mates with scores less than $u_k^*$. If we let $h(u)$ denote the probability density function of male quality scores $u$ in the local population, then the optimal rejection threshold for the female is defined by the equality

$$R(u_k^*) = P \cdot \frac{\int_{u_k^*}^{\infty} R(u)\, h(u)\, du}{\int_{u_k^*}^{\infty} h(u)\, du} + (1 - P) \cdot (-C) \tag{5}$$

On the left-hand side of this equation is the reward $R(u)$ attributed by the female to accepting a male with quality score $u = u_k^*$. For males with the critical quality score $u_k^*$ this reward is exactly equal to the expected reward after rejecting the male, which is on the right-hand side of the equation. Here, the first term reflects the expected reward of accepting a male with quality score $u > u_k^*$, multiplied by the probability $P$ that the female will encounter such a male before the reproductive season ends. The second term on the right-hand side reflects the cost ($C$) of not being able to produce an offspring, in the event that the female is unable to find an acceptable mate before end of season.

The probability of encountering an acceptable male, contingent on continuation of the reproductive season, is given by $p_k = \int_{u_k^*}^{\infty} h(u)\, du = 1 - H(u_k^*)$, where $H(u)$ is the cumulative distribution function of male quality scores. Mate search terminates either when the season ends, or when the reproductive season continues and the female encounters an acceptable mate. Accordingly, the probability $P$ is given by

$$P = \frac{p_k\,(1 - q)}{q + p_k\,(1 - q)} \tag{6}$$

From this expression it follows that females run a substantial risk of remaining unmated, unless $p_k \gg q$.

Let us now assume that the function $S(\Delta x, y)$ defined in Eq. (4) expresses male quality scores relative to the (fitness) opportunity cost of rejecting a male. Then, the reward function takes the form $R(u) = u q C$. After substituting this definition in Eq. (5), simplifying and ignoring small terms of order $q$, we obtain a simple condition for the threshold of mate acceptance:

$$\int_{u_k^*}^{\infty} (u - u_k^*) h(u) \, du = 1 \tag{7}$$

According this result, the female's threshold for mate acceptance $u_k^*$ is such that the potential improvement in mate quality that can be realised by searching for a better candidate is exactly compensated by the risk of remaining unmated.

The distribution of male scores is not fixed, due to the dynamics of ecological and mating traits. Therefore, we assume that females infer the local distribution $h(u)$ based on their own phenotype and on the sample of males $M = \{m_1, m_2, \ldots, m_n\}$ so far encountered during the mating season. The same information is used by the female to estimate $\sigma$, the standard deviation of the ecological trait. Accordingly, when encountering the $n$-th male, the female assigns the score

$$u_{k \times m_n} = -a \, y_k \frac{(x_k - x_{m_n})^2}{\frac{1}{n} \left((x_k - \bar{x})^2 + \sum_{m \in M}(x_m - \bar{x})^2\right)} = -a \, y_k \frac{(x_k - x_{m_n})^2}{\frac{1}{n}\left(\sum_{m \in M}(x_k - x_m)^2 - \frac{(n \, x_k - \sum_{m \in M} x_m)^2}{n+1}\right)} \tag{8}$$

and her threshold for mate rejection is determined by integrating over the discrete distribution of quality scores observed so far. This integral is computed by applying partial integration to the left-hand side of Eq. (7), converting to a discrete sum and simplifying using summation by parts. For this calculation, it is convenient to order the males $m_1, m_2, \ldots, m_n$ in the sample $M$ according to their quality score such that $u_{m_1} \geqslant u_{m_2} \geqslant \ldots \geqslant u_{m_n}$. Then, the discretized equivalent of Eq. (7) can be obtained as follows:

$$\begin{aligned}
\int_{u_k^*}^{\infty} (u - u_k^*) h(u) \, du &= \int_{u_k^*}^{\infty} (1 - H(u)) \, du \\
&= \frac{L}{n} (u_{m_L} - u_k^*) + \sum_{1 \leqslant l < L} \frac{l}{n} (u_{m_l} - u_{m_{l+1}}) \\
&= \frac{L}{n} \left(\frac{1}{L} \sum_{l=1}^{L} u_{m_l} - u_k^*\right) = 1
\end{aligned} \tag{9}$$

7

where the index $L$ is chosen such that $u_{m_L} \geqslant u_k^* \geqslant u_{m_{L+1}}$.

## Mapping from genotype to phenotype

### Gene-interactions

### Simulation data analysis

## Juvenile survival and population-density regulation

### Migration between patches

# Results

## subsection

## subsection

# Discussion

# Conclusion

# Acknowledgments

# Online Appendix

**A   Supplementary Figures**

[Figure A1 goes here.]

[Figure A2 goes here.]

# B   Additional Methods

90

## Literature Cited

[93] Cook, O. E., G. H. Collaborator, and A. Q. Expert. 2015. Data from: Template and Guidelines for Using LaTeX in *The American Naturalist*. American Naturalist, Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.XYZAB.

[96] Davis, E. B., K. A. Brakora, and A. H. Lee. 2011. Evolution of ruminant headgear: a review. Proceedings of the Royal Society B 278:2857–2865.

Inglis, R. F., P. G. Roberts, A. Gardner, and A. Buckling. 2011. Spite and the scale of competition [99] in *Pseudomonas aeruginosa*. American Naturalist 178:276–285.

Lemodèle, P.-Q., A. B. Kapitelschreiber, and C. D. E. Exemplar. 2007. An exemplary instance of chapters in books. Pages 231–245 *in* J.-P. Écrivain and M. A. Természettudós, eds. Inspiring [102] Instances of Brilliant Writing. Truth Pudding Press, Fond du Lac, WI.

Xiao, X., D. J. McGlinn, and E. P. White. 2015. A strong test of the maximum entropy theory of ecology. American Naturalist 185:E705–E80.

# Tables

Table 1: Animals in various cities with equations

| Animal | City | Equation |
|---|---|---|
| Dog | Springfield | $x + y = z$ |
| Fox | Indianapolis | $2x + 2y = 2z$ |
| Okapi[a] | Chicago | $x - y < z$ |
| Badger | Madison | $x + 2y > z$ |

Note: Table titles should be short. Further details should go in a 'notes' area after the tabular environment, like this.

[a] Okapis are not native to Chicago, but they are to be met with in both of the major Chicagoland zoos.

## Figure legends

Figure 1: Figure legends can be longer than the titles of tables. However, they should not be excessively long.

Video 1: Video legends can follow the same principles as figure legends. Counters should be set and reset so that videos and figures are enumerated separately.

Figure 2: In this way, figure legends can be listed at the end of the document, with references that work, even though the graphic itself should be included for final files after acceptance. Instead, upload the relevant figure files separately to Editorial Manager; Editorial Manager should insert them at the end of the PDF automatically.

### Online figure legends

Figure A1: *A*, the quick red fox proceeding to jump 20 m straight into the air over not one, but several lazy dogs. *B*, the quick red fox landing gracefully despite the skepticism of naysayers.

Figure A2: The quicker the red fox jumps, the likelier it is to land near an okapi. For further details, see Lemodèle et al. (2007).