Spiraling Complexity: A Test of the Snowball Effect in a Computational Model of RNA Folding

Ata Kalirad* **and Ricardo B. R. Azevedo***,[1]
*Department of Biology and Biochemistry, University of Houston, Houston, TX, US

ABSTRACT Genetic incompatibilities can emerge as a by-product of genetic divergence. According to Dobzhansky and Muller, an allele that fixes in one population may be incompatible with an allele at a different locus in another population when the two alleles are brought together in hybrids. Orr showed that the number of Dobzhansky–Muller incompatibilities (DMIs) should accumulate faster than linearly—i.e., snowball—as two lineages diverge. Several studies have attempted to test the snowball effect using data from natural populations. One limitation of these studies is that they have focused on predictions of the Orr model but not on its underlying assumptions. Here we use a computational model of RNA folding to test both predictions and assumptions of the Orr model. Two populations are allowed to evolve in allopatry on a holey fitness landscape. We find that the number of DMIs involving pairs of loci (i.e., simple DMIs) does not snowball—rather, it increases approximately linearly with divergence. We show that the probability of emergence of a simple DMI is approximately constant, as assumed by the Orr model. However, simple DMIs can disappear after they have arisen, contrary to the assumptions of the Orr model. This occurs because simple DMIs become complex (i.e., involve alleles at three or more loci) as a result of later substitutions. We introduce a modified Orr model where simple DMIs can become complex after they appear. Our modified Orr model can account for the results of the RNA folding model. We also find that complex DMIs are common and, unlike simple ones, do snowball. Reproductive isolation, however, does not snowball because DMIs do not act independently of each other. We conclude that the RNA folding model supports the central prediction of the Orr model that the total number of DMIs snowballs, but challenges some of its underlying assumptions.

KEYWORDS speciation; hybrid incompatibility; missing snowball; intrinsic postzygotic isolation; high-order epistasis

> "[It is not] surprising that the facility of effecting a first cross, the fertility of the hybrids produced, and the capacity of being grafted together … should all run, to a certain extent, parallel with the systematic affinity of the forms which are subjected to experiment …" Darwin (1859)

In the absence of gene flow, the gradual accumulation of divergent genetically based characteristics in different populations can bring new species into being. Some of these divergent characteristics, known as reproductive isolating barriers (Johnson 2006), decrease the level of interbreeding between populations. As populations diverge, isolating barriers accumulate, and the level of reproductive isolation (RI) among populations increases (Coyne and Orr 1989; Roberts and Cohan 1993; Sasa *et al.* 1998; Edmands 2002; Presgraves 2002; Lijtmaer *et al.* 2003; Mendelson 2003; Dettman *et al.* 2003; Moyle *et al.* 2004; Bolnick and Near 2005; Liti *et al.* 2006; Scopece *et al.* 2007; Stelkens *et al.* 2010; Jewell *et al.* 2012; Giraud and Gourbière 2012; Larcombe *et al.* 2015). Eventually RI reaches a point where two of these populations are considered distinct species. Elucidating the precise nature of the relationship between divergence and RI remains one of the central challenges in the study of speciation (Gavrilets 2004; The Marie Curie SPECIATION Network 2012; Nosil and Feder 2012; Seehausen *et al.* 2014).

Dobzhansky (1937) and Muller (1942) proposed a general

Copyright © 2016 by the Genetics Society of America
doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Wednesday 21st December, 2016%
[1]Corresponding author: Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5001. E-mail: razevedo@uh.edu

mechanism through which genetic divergence can cause RI. They noted that, in the absence of gene flow between two populations, an allele that fixes in one population may be incompatible with an allele at a different locus in another population when the two alleles are brought together in hybrids. This negative epistasis, or genetic incompatibility, causes the two populations to become reproductively isolated. Dobzhansky-Muller incompatibilities (DMIs) have been shown to cause inviability or sterility in hybrids between closely related species (reviewed in Presgraves 2010b; Rieseberg and Blackman 2010; Maheshwari and Barbash 2011).

Orr (1995) modeled the accumulation of multiple DMIs as populations diverge. Consider two populations diverged at k loci and showing I_k simple DMIs. A *simple* DMI is defined as a negative epistatic interaction between an allele at one locus in one population and an allele at a different locus in the other population. Orr showed that when the next substitution takes place, the expected number of simple DMIs is

$$I_{k+1} = I_k + kp \quad , \tag{1}$$

where p is the probability that there is a simple DMI between the latest derived allele and one of the k alleles at the loci that have previously undergone substitutions (from the population that did not undergo the latest substitution). Assuming $I_1 = 0$, the solution to Equation 1 is

$$I_k = \frac{k(k-1)p}{2} \quad . \tag{2}$$

Equation 2 predicts that the number of simple DMIs will accumulate faster than linearly as a function of divergence (prediction #1; Orr 1995). This prediction assumes that p remains constant as populations diverge (assumption #1).

DMIs involving three or more loci, known as *complex* DMIs (Cabot *et al.* 1994), are also expected to snowball but following different relationships from that in Equation 2: DMIs of order n are expected to accumulate at a rate approximately proportional to k^n (prediction #2; Orr 1995; Welch 2004). If DMIs have small, independent effects on RI (assumptions #2 and #3, respectively), then the postzygotic RI they generate is also expected to increase faster than linearly with k (prediction #3; Orr 1995). Orr (1995) described these patterns of quantities increasing faster than linearly as "snowballing." We shall refer to predictions #1–3 of the Orr model collectively as the "snowball effect" (Orr and Turelli 2001).

Several studies have attempted to test the snowball effect. They have employed three different approaches. The first tests prediction #3 of the Orr model: that postzygotic RI snowballs. For example, Larcombe *et al.* (2015) measured the strength of hybrid incompatibility between *Eucalyptus globulus* and 64 species of eucalypts. They observed a faster than linear increase in RI with genetic distance, consistent with prediction #2 of the Orr model. Results from other studies using a similar approach have provided little support for a snowball effect in RI (Sasa *et al.* 1998; Lijtmaer *et al.* 2003; Mendelson *et al.* 2004; Bolnick and Near 2005; Gourbière and Mallet 2010; Stelkens *et al.* 2010; Giraud and Gourbière 2012), leading some to pronounce the snowball "missing" (Johnson 2006; Gourbière and Mallet 2010). However, this approach has several limitations. It can only be applied when postzygotic RI $\ll 1$. Furthermore, it only tests one prediction (#3) of the Orr model, and this prediction relies on one assumption (#3) that typically goes untested. Thus, the number of DMIs might snowball (predictions #1–2 of) even if RI does not.

The second approach tests predictions #1–2 of the Orr model: that the number of DMIs snowballs. For example, Moyle and Nakazato (2010) used a QTL mapping approach to estimate the number of DMIs between species of *Solanum* directly. They introgressed one or a few genomic segments from one species to another. When an introgressed segment caused a reduction in fitness, they concluded that it participated in a DMI. They found that DMIs affecting seed sterility accumulated faster than linearly. However, DMIs affecting pollen sterility appeared to accumulate linearly, contrary to the snowball effect. Studies following similar approaches have tended to find support for the snowball effect (Matute *et al.* 2010; Moyle and Nakazato 2010; Matute and Gavin-Smyth 2014; Sherman *et al.* 2014; Wang *et al.* 2015). One advantage of the second approach over the first is that it relies on fewer assumptions (#1 compared to #1–3, respectively). However, the second approach also has limitations. The order (n) of the DMIs identified is unknown. Therefore, this approach cannot disentangle predictions #1 and #2. Another limitation of these studies is that they are likely to underestimate the true number of DMIs for two reasons. First, the introgressed genomic segments typically contain many genetic differences. For example, the individual segments introgressed in Moyle and Nakazato (2010) included approximately 2–4% of the genome, and likely contained hundreds of genes. Second, individual alleles might participate in multiple DMIs, specially if complex DMIs are common.

The third approach tests prediction #1 of the Orr model: that the number of simple DMIs snowballs. Consider two species, 1 and 2, diverged at k loci. If an allele, X_2, at one of these loci (X) is known to be deleterious in species 1 but is fixed in species 2, then species 2 must carry compensatory alleles at one or more loci $(Y_2, Z_2, \ldots)$ that are not present in species 1 (which carries alleles $Y_1, Z_1, \ldots$ at those loci). In other words, there must be a DMI involving the X_2 and $Y_1, Z_1, \ldots$ alleles.

Following Welch (2004), we define $\mathcal{P}_1$ as the proportion of the k fixed differences between the species where the allele from one species is deleterious in the other species. For example, Kachroo *et al.* (2015) replaced 414 essential genes of the yeast *Saccharomyces cerevisiae* with their human orthologs. Over half of the human genes $(\mathcal{P}_1 = 57\%)$ could not functionally replace their yeast counterparts.

Welch (2004) has argued that estimates of $\mathcal{P}_1$ can be used to test the Orr model if two additional conditions are met. If each allele participates in at most one DMI, then we have $\mathcal{P}_1 = I_k/k$. If, in addition, $\mathcal{P}_1$ is entirely based on simple DMIs, then it is expected to increase linearly with genetic distance according to the Orr model (Equation 2)

$$\mathcal{P}_1 = \frac{(k-1)p}{2} \quad . \tag{3}$$

Interestingly, $\mathcal{P}_1$ can be estimated without studying hybrids directly. Kondrashov *et al.* (2002) and Kulathinal *et al.* (2004) estimated $\mathcal{P}_1$ in mammals and insects, respectively. Surprisingly, both studies reported that $\mathcal{P}_1 \approx 10\%$ and is constant over broad ranges of genetic distances (e.g., human compared to either nonhuman primates or fishes, Kondrashov *et al.* 2002). These results are inconsistent with prediction #1 of the Orr model (Welch 2004; Fraïsse *et al.* 2016). The results of the second and third approaches give inconsistent results, a paradox first noted by Welch (2004). However, the third approach is less direct because it relies on two additional assumptions that have not been tested.

One common limitation to all approaches is that they focus on testing predictions of the Orr model, without testing its assumptions (e.g., assumption #1, constant p). Here we use a computational model of RNA folding (Schuster *et al.* 1994; Lorenz *et al.* 2011) to test both predictions and assumptions of the Orr model. The RNA folding model makes satisfactory predictions of the secondary structures of real RNA molecules (Mathews *et al.* 1999; Doshi *et al.* 2004; Lorenz *et al.* 2011) and has been used to study other evolutionary consequences of epistasis, including robustness (van Nimwegen *et al.* 1999; Ancel and Fontana 2000), evolvability (Wagner 2008; Draghi *et al.* 2010), and the rate of neutral substitution (Draghi *et al.* 2011). We model populations evolving in allopatry on a holey fitness landscape (Gavrilets 2004). In his original model, Orr (1995) made no assumptions on either the evolutionary causes of genetic divergence, or the molecular basis of the DMIs arising from this divergence. Thus, Orr's predictions should be met in our RNA "world." Our results provide mixed support for the Orr model.

Materials and Methods

Genotype and phenotype

The genotype is an RNA sequence. Unless otherwise stated we used sequences with a length of 100 nucleotides. The phenotype is the minimum free-energy secondary structure of the sequence computed using the ViennaRNA package 2.1.9 (Lorenz *et al.* 2011) with default parameters.

Fitness

The fitness of RNA sequence i is determined using the step function:

$$w_i = \begin{cases} 1 & \text{if} \quad \beta_i > \alpha \quad \text{and} \quad \delta_i \leqslant \alpha \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where β_i is the number of base pairs in the secondary structure of sequence i, δ_i is the base-pair distance between the structure of sequence i and the reference structure, and α is an arbitrary threshold. Unless otherwise stated we used $\alpha = 12$. The fitness function in Equation 4 specifies a neutral network (Schuster *et al.* 1994; van Nimwegen *et al.* 1999), a type of holey fitness landscape (Gavrilets 2004) (Figure 1).

Evolution

Burn-in period: We begin by picking a random viable RNA sequence, define its secondary structure as the reference, and allow it to accumulate 200 random neutral substitutions sequentially, allowing multiple hits. The resulting sequence is used as the ancestor. Table S1 shows summary statistics for the ancestral sequences for $\alpha = 12$.

The burn-in period is necessary because the initial sequence is not representative for the fitness landscape. For example, it has the reference structure (i.e., $\delta_i = 0$ base pairs), whereas most sequences in the fitness landscape are $\delta_i \approx \alpha$ base pairs away from the reference structure (Table S1).

Divergence: The ancestor is used to found two identical haploid lineages. The lineages evolve by alternately accumulating a series of neutral substitutions without gene flow (allopatry) until they differ at $k = 40$ sites. At a given step, one of the evolving sequences is subjected to a random mutation. If the mutation is neutral, it is allowed to substitute; if it is deleterious, it is discarded and a new random mutation is tried. The process is repeated until a neutral mutation is found. At the next step, the other evolving lineage is subjected to the same process.

At each step, the only sites that are allowed to mutate are those that have not yet undergone a substitution in either lineage since the lineages have started to diverge from their common ancestor. This constraint implies that no more than two alleles are observed at each site during the course of evolution and that substitutions are irreversible, in agreement with the assumptions of the Orr (1995) model. All types of base-substitution mutations have equal probability. Insertions and deletions are not considered.

Detecting DMIs

In this section we use the general terms genotypes, loci and alleles, instead of sequences, sites and nucleotides.

Two genotypes, 1 and 2, both have fitness $w = 1$ and differ at $k \geq 2$ loci. Loci are denoted by $A, B, C, \ldots$ The alleles of genotype 1 are indicated by a subscript 1 $(A_1, B_1, C_1, \ldots)$; the alleles of genotype 2 are indicated by a subscript 2 $(A_2, B_2, C_2, \ldots)$. Introgression of the A_1 and B_1 alleles from genotype 1 to genotype 2 is denoted $1 \xrightarrow{A,B} 2$.

Simple DMIs: There is a simple DMI between the A_1 and B_2 alleles if *all* of the following 6 conditions are met.

1. The single introgression $1 \xrightarrow{A} 2$ results in an inviable genotype (Figure 2, step I). On its own, this condition indicates that there is a DMI between the A_1 allele and one or more alleles from genotype 2 at the remaining $k-1$ loci $(B_2, C_2, \ldots)$.

2. The single introgression $2 \xrightarrow{B} 1$ results in an inviable genotype. On its own, this condition indicates that there is a DMI between the B_2 allele and one or more alleles from genotype 1 at the remaining $k-1$ loci $(A_1, C_1, \ldots)$. Taken together, conditions #1–2 are not sufficient to indicate that the A_1 and B_2 alleles participate in the same DMI.

3. The double introgressions $1 \xrightarrow{A,B} 2$ and $2 \xrightarrow{A,B} 1$ both result in viable genotypes (Figure 2, step II). In other words, a second introgression rescues viability. Taken together, conditions #1–3 indicate that the A_1 and B_2 alleles participate in the same DMI; the conditions do not, however, rule out the possibility that the DMI involves additional alleles from either genotype at the remaining $k-2$ loci $(C, D, \ldots)$. In other words, the DMI might be simple or complex.

4. A_1 and B_2 are not both ancestral (Orr 1995). If conditions #1–3 are met but condition #4 is violated, then the DMI must involve a derived allele at an additional locus—i.e., the DMI is complex—because A_1 and B_2 were not incompatible in the ancestor.

5. If both A_1 and B_2 are derived alleles, this condition is ignored. If A_1 is an ancestral allele, then the B_2 substitution occurred after the A_2 substitution; if B_2 is an ancestral allele, then the A_1 substitution occurred after the B_1 substitution (Orr 1995). If conditions #1–4 are met but condition #5 is violated then the DMI is complex because A_1 and B_2 were not incompatible in the background in which the derived allele arose.

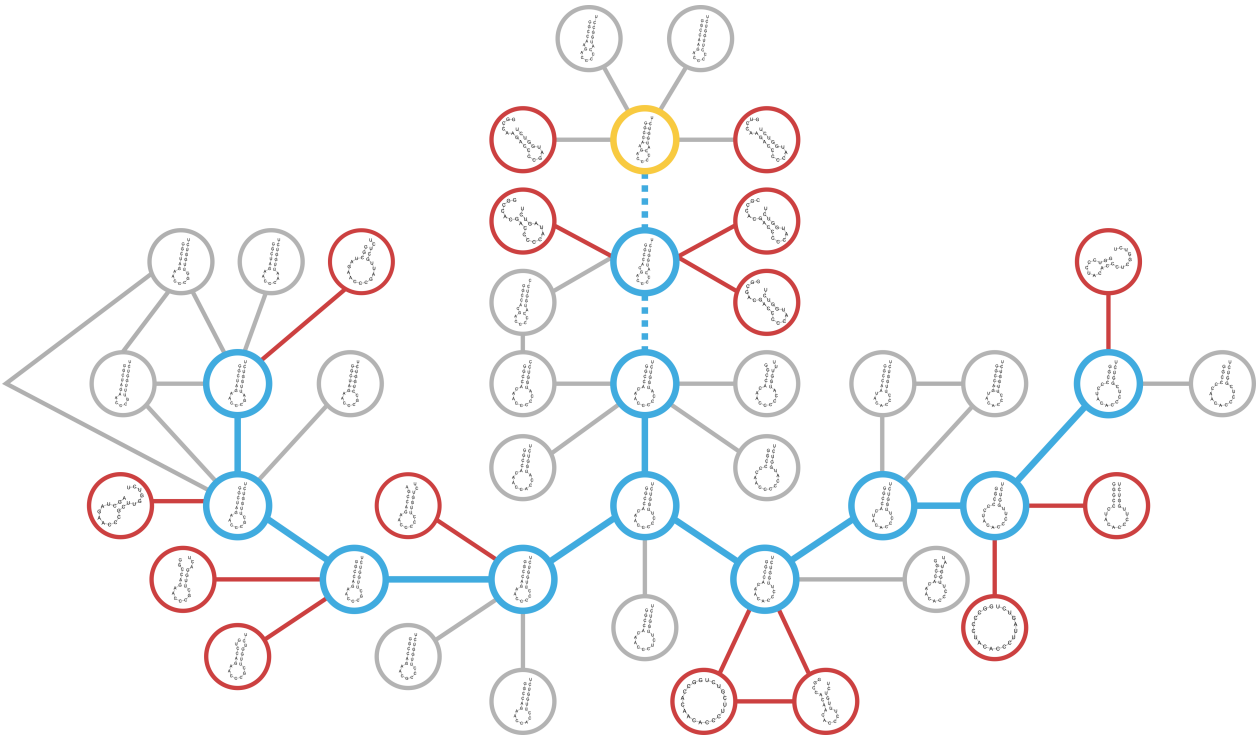

Figure 1 Evolution on a holey fitness landscape. Mutational network of RNA sequences. Lines connect sequences of 20 nucleotides that can be reached by a single nucleotide substitution. Only a tiny fraction of the entire mutational network of $\sim 10^{12}$ sequences is shown. Furthermore, only a few of the 60 mutational neighbors of each sequence are shown. A sequence is viable (yellow, blue or gray circles) if its secondary structure both has more than $\alpha = 2$ base pairs and is at most $\alpha = 2$ base pairs away from the reference structure (thick yellow circle); a sequence is inviable otherwise (red circles) (Equation 4). Each simulation starts with a burn-in period where a sequence with the reference structure undergoes 3 neutral substitutions (thick dashed blue lines). After that, the resulting sequence is used as the ancestor of two lineages that alternately accumulate neutral substitutions until they have diverged at $k = 8$ sites (thick solid blue lines).

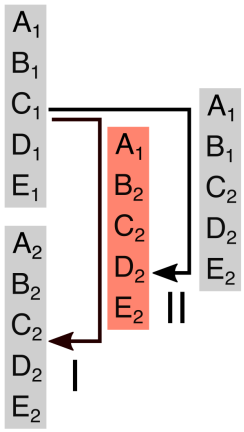

Figure 2 Detecting DMIs. To find simple DMIs, we use an introgression–rescue assay where we introgress one diverged allele between the two lineages (step I: $1 \xrightarrow{A} 2$), and if this substitution results in an inviable genotype (red), we try to rescue it with a second introgression (step II: $1 \xrightarrow{A,B} 2$). If the second introgression rescues viability, we conclude that the there is a DMI between the first introgressed allele (A_1) and the resident allele at the second locus (B_2). The additional criteria for establishing whether the DMI is simple or complex are explained in the Materials and Methods.

6. If the latest substitution at either the A or the B locus was the i-th substitution, and $i < k$, then conditions #1–3 are also met in the genotypes present immediately after the i-th substitution. If conditions #1–5 are met but condition #6 is violated then the DMI is complex because A_1 and B_2 were not incompatible in the background in which the latest derived allele arose.

Complex DMIs: Imagine that condition #1 for a simple DMI is met: a single introgression $1 \xrightarrow{A} 2$ results in an inviable genotype. As explained above, this is indicative of a DMI involving the A_1 allele. This DMI is complex if *any* of the following 4 conditions are met.

7. It satisfies conditions #2–3 for a simple DMI but violates one or more of conditions #4–6.

8. The double introgression $1 \xrightarrow{A,B} 2$ rescues viability, but the single introgression $2 \xrightarrow{B} 1$ results in a viable genotype (i.e., condition #2 is violated).

9. The double introgression $1 \xrightarrow{A,B} 2$ rescues viability, but the double introgression $2 \xrightarrow{A,B} 1$ results in an inviable genotype (i.e., condition #3 is violated).

10. There is no double introgression of the form $1 \xrightarrow{A,B} 2$ that rescues viability (i.e., condition #3 is violated).

A DMI is also complex if it satisfies the following condition:

11. The introgression of $1 < i < k$ alleles (e.g., $1 \xrightarrow{A,B,\ldots} 2$) results in an inviable genotype, but all the introgressions of each individual allele and of any combination of between 2 and $i-1$ of the alleles result in a viable genotype. This condition indicates that the i alleles participate in a complex DMI of order $n \geqslant i+1$.

Assays

Number of simple DMIs: To count simple DMIs in our simulations, we introgress nucleotides between the two sequences at each of the k divergent sites, in both directions. Every time an introgression results in an inviable genotype (condition #1), we look for another introgression in the opposite direction that also results in an inviable genotype (condition #2). We then test both double introgressions involving these alleles to test for condition #3. If we find a pair of alleles satisfying conditions #1–3, we test for conditions #4–6 directly. We count simple DMIs after every substitution when $k \geqslant 2$.

Number of complex DMIs: The criteria described above (conditions #7–11) allow us to detect complex DMIs. However, counting them for highly diverged sequences (high k) is virtually impossible for two reasons. First, the number of high-order introgressions required is enormous. Second, as the conditions #1–3 for detecting simple DMIs highlight, establishing that alleles participate in the *same* DMI requires additional introgressions. For example, if alleles A_1 and B_1 from population 1 are incompatible with allele C_2 from population 2, then both the double introgression $1 \xrightarrow{A,B} 2$ and the single introgression $2 \xrightarrow{C} 1$ result in an inviable genotype. However, showing that the 3 alleles are involved in the same DMI of order $n = 3$ would require demonstrating that the triple introgressions $1 \xrightarrow{A,B,C} 2$ and $2 \xrightarrow{A,B,C} 1$ both result in viable genotypes.

Thus, without conducting "rescue" introgressions, the introgressions in both directions will tend to overestimate the number of complex DMIs. To avoid this problem, we estimate the number of complex DMIs through all single, double and triple introgressions in one direction only (e.g., from population 1 to population 2). For the single introgressions, we count complex DMIs using conditions #7–10 (these conditions require performing introgressions in both directions, but only DMIs detected from an introgression in one direction are counted). For double and triple introgressions, we use condition #11.

The resulting count of complex DMIs will still underestimate the true number for two reasons. First, if the introgressed alleles participate in more than one complex DMI, an introgression test can only detect a single DMI (this limitation does not apply to simple DMIs). Second, complex DMIs that can only be detected by introgressing four or more alleles will not be detected.

Proportion of single introgressions involved in a DMI: We use the single introgression data to calculate $\mathcal{P}_1$, the proportion of the $2k$ single introgressions at diverged sites (in both directions) that result in an inviable sequence (Welch 2004). Both simple and complex DMIs are expected to influence $\mathcal{P}_1$.

DMI network: The simple DMIs that might, potentially, affect a sequence can be computed exhaustively by measuring the fitness of all possible single and double mutants derived from the sequence. For every pair of sites, there are 9 combinations of double mutants. A potential simple DMI is defined as an inviable double mutant between mutations that are individually neutral. We summarize the pattern of interactions between sites using an undirected network where the vertices are sites and the edges represent the existence of at least one potential simple DMI between them. The resulting network is an example of the networks of interactions described by Orr and Turelli (2001) and Livingstone *et al.* (2012).

We measure the degree of similarity between two DMI networks X and Y using the Jaccard index

$$J = \frac{|X \cap Y|}{|X \cup Y|} \quad , \tag{5}$$

where $|X \cap Y|$ is the number of edges shared between the two networks, $|X \cup Y|$ is the sum of $|X \cap Y|$ and the numbers of edges unique to X and Y, and there is a one-to-one correspondence between the vertices of X and Y (i.e., between the sites in the corresponding sequences). J varies between 0 (the two networks have no edges in common) and 1 (the two networks are identical).

Reproductive isolation: The degree of RI between two sequences is defined as

$$\mathrm{RI} = 1 - \overline{w}_R \quad ,$$

where $\overline{w}_R$ is the mean fitness (Equation 4) of all possible 198 recombinants resulting from a single crossover between the sequences.

"Holeyness" of the fitness landscape: For each simulation, we took the ancestor and each of the $k = 40$ genotypes generated during the course of evolution and measured the proportion of their single mutant neighbors (300 per sequence) that are inviable, excluding the 41 original sequences. This estimates the local holeyness of the fitness landscape traversed by the diverging lineages.

Direct simulation of the Orr model

We also simulate the accumulation of DMIs following the Orr (1995) model. An ancestral genotype has multiple loci and is used to found two identical haploid lineages. The lineages are allowed to evolve by alternately accumulating neutral substitutions (Figure 3).

	Lineage 1	Lineage 2
$k = 0$	$A_0\, B_0\, C_0\, D_0 \ldots$	$A_0\, B_0\, C_0\, D_0 \ldots$
$k = 1$	$\mathbf{A_1}\, B_0\, C_0\, D_0 \ldots$	$A_0\, B_0\, C_0\, D_0 \ldots$
$k = 2$	$\mathbf{A_1}\, B_0\, C_0\, D_0 \ldots$	$A_0\, \mathbf{B_2}\, C_0\, D_0 \ldots$
$k = 3$	$\mathbf{A_1}\, B_0\, \mathbf{C_1}\, D_0 \ldots$	$A_0\, \mathbf{B_2}\, C_0\, D_0 \ldots$
$k = 4$	$\mathbf{A_1}\, B_0\, \mathbf{C_1}\, D_0 \ldots$	$A_0\, \mathbf{B_2}\, C_0\, \mathbf{D_2} \ldots$

Figure 3 Sequence evolution in a direct simulation of the Orr model showing the first $k = 4$ substitutions. Only 4 loci are shown, denoted by A–D. Ancestral alleles are indicated by subscript 0. Derived alleles are shown in bold and indicated by subscripts 1 or 2 depending on the lineage.

After the k-th substitution, simple DMIs are sampled at random with probability p_k from all pairs of alleles consisting of the latest derived allele paired with any of the $k - 1$ ancestral or derived alleles from the other population at loci that have previously undergone substitutions in either population. For example, when $k = 4$ the new possible simple DMIs are: D_2/A_1, D_2/B_0, and D_2/C_1 (Figure 3).

Statistical analyses

All statistical analyses were conducted with R version 3.3.0 (R Core Team 2016). Partial rank correlations were calculated using the "ppcor" package (Kim 2015).

Data availability

The software used to run all simulations was written in Python 2.7 and is available at https://github.com/Kalirad/spiraling_complexity (DOI: https://doi.org/10.5281/zenodo.216743). The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

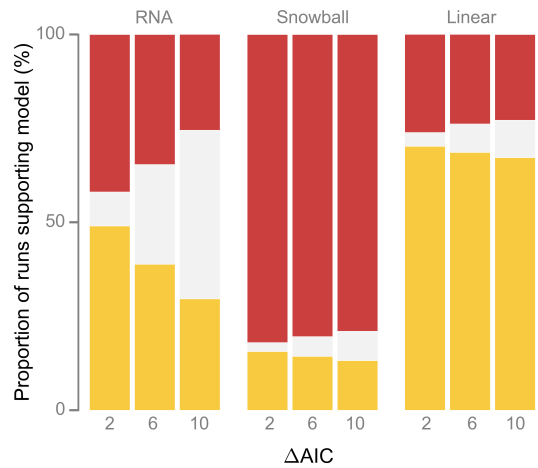

Figure 4 Simple DMIs do not snowball in the RNA folding model. We fitted the snowball model (Equation 2) and a linear model (Equation 6) to each run from three kinds of simulations: simulations of the RNA folding model ("RNA"), direct simulations of the Orr model ("Snowball") with values of p estimated by fitting the model in Equation 2 to each RNA folding simulation (Figure S1A), and direct simulations of the Orr model ("Linear") with values of p_k from Equation 7 estimated by fitting the model in Equation 6 to each RNA folding simulation (Figure S1B). Red segments show the proportions of runs providing stronger support for the snowball model; yellow segments show the proportions of runs providing stronger support for the linear model; gray segments show the proportions of runs providing approximately equal support for both models. Each bar is based on 10^3 stochastic simulations. The level of support for the two models was evaluated for three different ΔAIC thresholds.

Results

Simple DMIs do not snowball in the RNA folding model

The Orr (1995) model predicts that the number of simple DMIs, I_k, should increase faster than linearly with the number of substitutions, k. We tested this prediction (#1) using 10^3 evolutionary simulations with the RNA folding model. For each simulation, we fitted two models: the snowball model in Equation 2 and a linear model of the form

$$I_k = (k - 1)b \quad , \tag{6}$$

where b is the slope. The $k - 1$ term ensures that $I_1 = 0$, as in the snowball model. Both models have a single parameter that we estimated using the method of least squares.

We compared the level of support for each model using Akaike's Information Criterion (AIC). If the difference in the AIC values (ΔAIC) was greater than a threshold, we concluded that there was stronger support for the model with the lower AIC. Setting the ΔAIC threshold at 2, 41.9% of RNA folding

simulations provided stronger support for the snowball model, 49.1% provided stronger support for the linear model, and 9.0% provided approximately equal support for both models (Figure 4). Increasing the ΔAIC threshold did not affect this result qualitatively (Figure 4).

To evaluate the extent to which the lack of support for the snowball model was caused by random noise in the simulations we conducted 10^3 direct simulations of a snowball process and 10^3 direct simulations of a linear process. As expected, snowball simulations provided stronger support for the snowball model and linear simulations provided stronger support for the linear model (Figure 4). RNA folding simulations showed a pattern of relative support for the snowball and linear models that more closely resembled that of the direct simulations of a linear process (Figure 4).

The number of simple DMIs in the RNA folding simulations accumulated at a rate approximately proportional to $k^{1.3}$ (Figure 5A; Table S2, c_2). Thus, it was more closely approximated by the linear model than by the snowball model, in agreement with the AIC analysis. We conclude that prediction #1 of the Orr model was not met in most RNA folding simulations.

The probability that a simple DMI appears is approximately constant in the RNA folding model

What explains the lack of support for a snowball effect in simple DMIs in the RNA folding simulations? One possibility is that p itself evolved, contrary to assumption #1 of the Orr model.

If p declines with divergence according to the relationship

$$p_k = \frac{b}{k} \quad , \tag{7}$$

where b is a positive constant, and we substitute p by p_k in Equation 1, the linear model in Equation 6 is a solution to the resulting difference equation (assuming $I_1 = 0$). To test whether p changed as described by Equation 7, we measured it directly in each simulation as $p_k = \Delta I / k$, where ΔI is the number of new simple DMIs appearing as a result of the $(k+1)$-th substitution that involve the latest derived allele (see Equation 1). We found that, although p_k declined with k, the trend did not follow Equation 7. Indeed, when $k \gtrsim 10$, p_k was approximately constant (Figure 5B).

Simple DMIs do not persist indefinitely in the RNA folding model

The previous analysis also revealed that fitting the snowball model to the RNA folding data underestimated the true value of p by approximately 3-fold (Figure 5B). This discrepancy indicates that a more fundamental assumption of the Orr model may be violated in the RNA folding model: that simple DMIs, once they have arisen, persist indefinitely (assumption #4). This assumption was not stated explicitly by Orr (1995) and has never, to our knowledge, been called into question.

To test assumption #4, we estimated the DMI networks of sequences as they evolved in our RNA folding model. Figure 6A shows an example of an RNA sequence evolving on a holey fitness landscape. Initially the sequence displays potential simple DMIs between 21 pairs of sites (Figure 6C). Figure 6B illustrates a potential simple DMI between positions 5 and 12. We refer to these simple DMIs as *potential* because if two diverging lineages each accumulate one of the substitutions underlying one of these DMIs, a simple DMI between the lineages will appear.

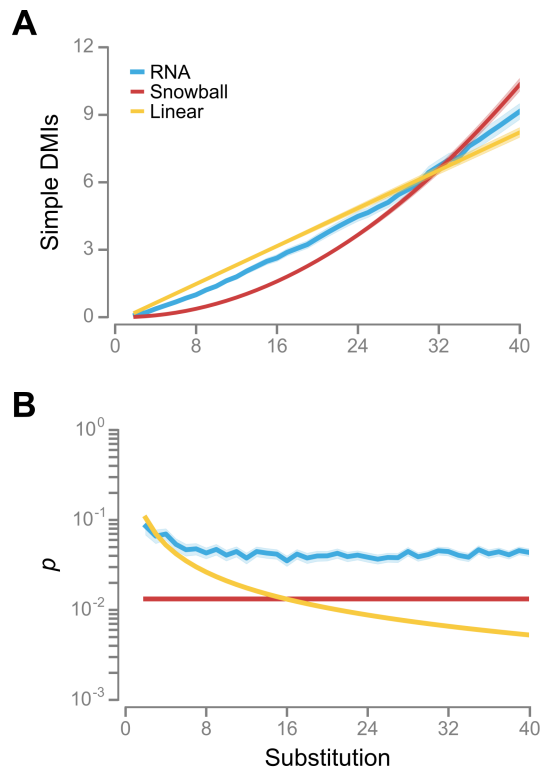

Figure 5 Simple DMIs do not snowball in the RNA folding model. (A) Evolution of the number of simple DMIs, I_k, as two populations diverge by accumulating substitutions, k. Values are means of 10^3 runs of three different kinds of stochastic simulations: "RNA," simulations of the RNA folding model (blue); "snowball," direct simulations of the Orr model with constant p_k estimated as explained in (B) (red); "linear," direct simulations of the Orr model with declining p_k estimated as explained in (B) (yellow). (B) Evolution of the probability, p_k, that there is a simple DMI between the latest derived allele after the $(k+1)$-th substitution and one of the k alleles at the loci that have previously undergone substitutions. The blue line ("RNA") shows the values of p_k estimated at each substitution directly from the RNA folding simulations. The red line ("Snowball") shows the values of p estimated by fitting the model in Equation 2 to each RNA folding simulation (Figure S1A). The yellow line ("Linear") shows the values of p_k from Equation 7 based on estimates of b obtained by fitting the model in Equation 6 to each RNA folding simulation (Figure S1B). Values are means of 10^3 simulations. Shaded regions indicate 95% confidence intervals, CIs.

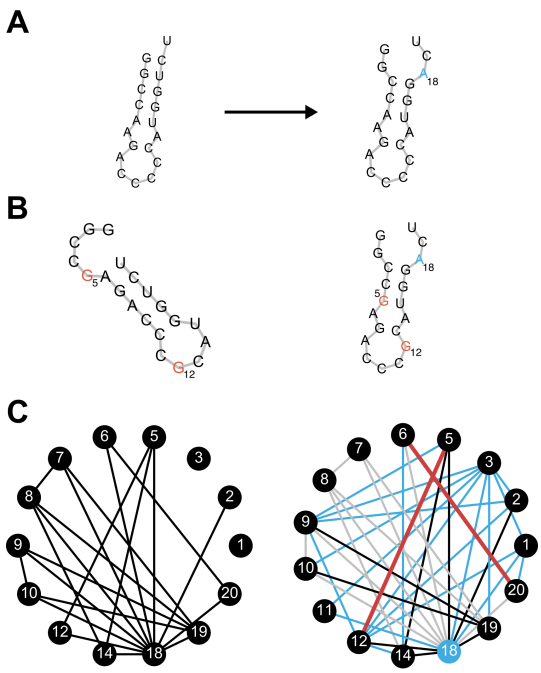

Figure 6 A single substitution can dramatically rearrange the network of potential DMIs. (A) The 20 nucleotide long RNA sequence on the left acquires a neutral U→A substitution at position 18 (blue). The holey fitness landscape is defined by $\alpha = 2$ (Equation 4). The secondary structure of the sequence on the left is the reference ($\delta_i = 0$ base pairs). The structure on the right is $\delta_i = 2$ base pairs away from the reference. (B) There is a potential simple DMI between positions 5 and 12 for the sequence on the left. A double mutant at those positions (5: A→G, 12: C→G, red) makes the structure inviable ($\delta_i = 11$ base pairs), even though the single mutations are neutral (not shown). However, a single substitution causes the potential simple DMI to disappear in the sequence on the right, although the single mutations remain neutral in the new background (not shown). In other words, the substitution causes the simple DMI to become complex. (C) DMI networks of the sequences in (A). Vertices correspond to positions in the sequences. An edge in the network on the left indicates that there is at least one potential simple DMI between the two sites (positions 4, 13 and 15–17 have no potential DMIs in either network and are not shown). Black edges in the network on the right are shared between the two networks. Blue edges exist only in the network on the right and indicate the appearance of new potential simple DMIs between sites caused by the substitution. Gray and red edges indicate disappearance of potential simple DMIs in the network on the right. Gray edges indicate disappearances due to the constituent alleles no longer being neutral in the new background. Red edges indicate disappearances caused by complexification; the DMI discussed in (B) is an example (5–12 edge). The Jaccard index (Equation 5) between the two networks is $J = 0.205$.

The Orr model assumes that the DMI network is static: as populations evolve they actualize potential DMIs (for an alternative, but equivalent, interpretation of DMI networks see Livingstone *et al.* 2012). However, DMI networks are not static in the RNA folding model. After a single neutral substitution, 13 pairs of sites (62%) lost all potential simple DMIs, and potential DMIs appeared between 18 new pairs of sites (Figure 6C).

The "disappearance" of a potential DMI can occur in one of two ways. First, the substitution may cause the mutations involved in the simple DMIs to become deleterious so that they can no longer participate in potential simple DMIs. A disappearance of this kind means that a potential simple DMI is no longer accessible through independent substitution in two lineages because one of the substitutions cannot take place. Thus, such disappearances do not contradict assumption #4 of the Orr model. The majority of disappearances in Figure 6C (gray lines) are of this kind.

The second kind of disappearance occurs when the substitution modifies the interaction between previously incompatible alleles (red lines in Figure 6C). In other words, the simple DMIs become complex. The potential simple DMI between positions 5 and 12 shown in Figure 6B disappears in this way. This kind of disappearance—complexification—implies that some simple DMIs may not persist indefinitely. In other words, assumption #4 is not always met in the RNA folding model.

The DMI networks corresponding to the evolving lineages in the RNA folding simulations summarized in Figure 5 also change dramatically relative to the ancestor as a result of successive substitutions (Figure S2). This indicates that complexification may be occurring in these simulations as well. In the next section we explore the consequences of the complexification of simple DMIs for the snowball effect.

The modified Orr model

We incorporate the dynamic nature of simple DMIs by extending the Orr (1995) model in Equation 1

$$I_{k+1} = (1-q)I_k + kp \quad , \tag{8}$$

where q is the probability that a simple DMI present after k substitutions becomes complex after the next substitution. Assuming $I_1 = 0$, the solution to Equation 8 is

$$I_k = \frac{p\left[(1-q)^k + kq - 1\right]}{q^2} \quad . \tag{9}$$

This prediction assumes that both p and q remain constant as populations diverge.

The original Orr model is a special case of the modified model when $q = 0$. When $q > 0$, the increase in the number of simple DMIs is given by

$$\Delta I = I_{k+1} - I_k = \frac{p}{q}\left[1 - (1-q)^k\right] \quad . \tag{10}$$

This equation has two consequences. First, the increase in the number of simple DMIs eventually becomes linear with a slope of approximately p/q when k is sufficiently large. Second, if q is larger, the "linearization" of Equation 9 occurs for lower values of k. Both patterns are illustrated in Figure 7A, which compares the accumulation of simple DMIs under the Orr model with $p = 0.04$ ($q = 0$), and that under the modified Orr model with the same value of p and increasing values of q.

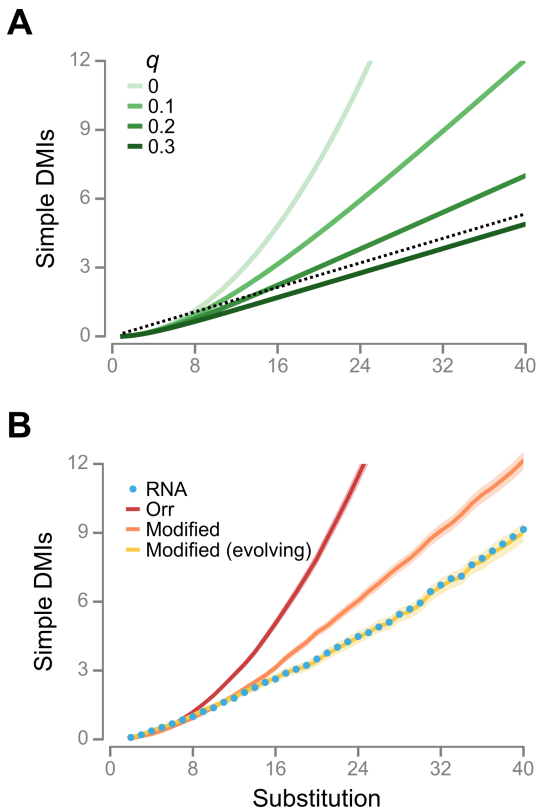

Figure 7 The RNA folding simulations agree with the modified Orr model. (A) Evolution of the number of simple DMIs under the modified Orr model. Responses for $p = 0.04$ and different values of q. The dashed line shows a slope of p/q for $q = 0.3$. (B) Mean responses of 10^3 runs of four different kinds of stochastic simulations: "RNA," simulations of the RNA folding model (blue circles, same data as in Figure 5A); "Orr," direct simulations of the Orr model with with constant values of p estimated directly from each RNA folding simulation (Figure 8) (red); "modified," direct simulations of the modified Orr model with constant values of p and q estimated directly from each RNA folding simulation (Figure 8) (orange); "modified (evolving)," direct simulations of the modified Orr model with evolving trajectories of p_k and q_k estimated directly from each RNA folding simulation (yellow, dashed). Shaded regions indicate 95% CIs.

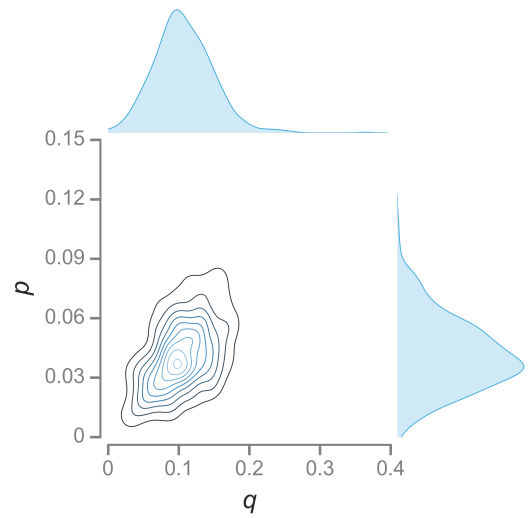

Figure 8 Distributions of the parameters of the modified Orr model in the RNA folding simulations: p, the probability that a simple DMI arises, and q, the probability that a simple DMI becomes complex. One- and two-dimensional kernel density estimates based on 10^3 stochastic simulations. For each simulation we calculated p_k and q_k after every substitution (k). We then estimated an overall value of p and q as weighted averages; values of p_k and q_k were weighted by $k(k-1)$ and I_k, respectively. The means of each distribution were $\bar{p} = 0.042$ and $\bar{q} = 0.107$.

The RNA folding simulations agree with the modified Orr model

To test whether the complexification of simple DMIs explains the results of the RNA folding simulations we measured q directly in our simulations as $q_k = 1 - I'_k/I_k$, where I_k is the number of simple DMIs present after the k-th substitution, and I'_k is the number of simple DMIs present after the $(k+1)$-th substitution that do not involve the latest derived allele.

The modified Orr model predicts that simple DMIs will accumulate approximately linearly when q is large relative to p (Equation 10). The values of q were, on average, 3-fold higher than the values of p (Figure 8). Furthermore, the q/p ratio was a good predictor of whether RNA folding simulations supported the linear or the snowball model (Figure 4). When the ΔAIC threshold was set at 2, q/p was 3.36 ± 0.22 (mean and 95% confidence intervals, CIs) in runs that provided stronger support for the linear model, and 2.41 ± 0.12 in runs that provided stronger support for the snowball model (Wilcoxon rank sum test, $P < 10^{-6}$). Thus, the approximately linear response in the number of simple DMIs in the RNA folding simulations can be explained by the modified Orr model.

To evaluate the extent to which the modified Orr model can account for the lack of support for a snowball effect in simple DMIs in our RNA folding simulations, we conducted 10^3 direct simulations of the modified Orr model over $k = 40$ substitutions assuming values of p and q estimated directly from the RNA folding data (Figure 8). The support for the snowball and linear models provided by these direct simulations of the modified Orr model was similar to that provided by the RNA folding simulations (Figure S3). These results, in combination with those on the q/p ratio, indicate that the complexification of

simple DMIs explains the RNA folding results.

Figure 7B shows that the modified Orr model (orange) approximates the RNA folding data better than the Orr model (red). However, the fit is far from perfect. The lack of fit is caused by the assumptions that both p and q are constant as populations diverge. Neither assumption was strictly met by the RNA folding data: p decreased and q increased with k, specially when $k \lesssim 10$ (Figures 5B and S4, respectively). When we allowed p and q to vary as they did in the RNA folding simulations, direct simulations of the modified Orr model matched the RNA folding data perfectly (Figure 7B). We conclude that the modified Orr model explains the RNA folding results for simple DMIs, provided we relax the assumptions that p and q are constant.

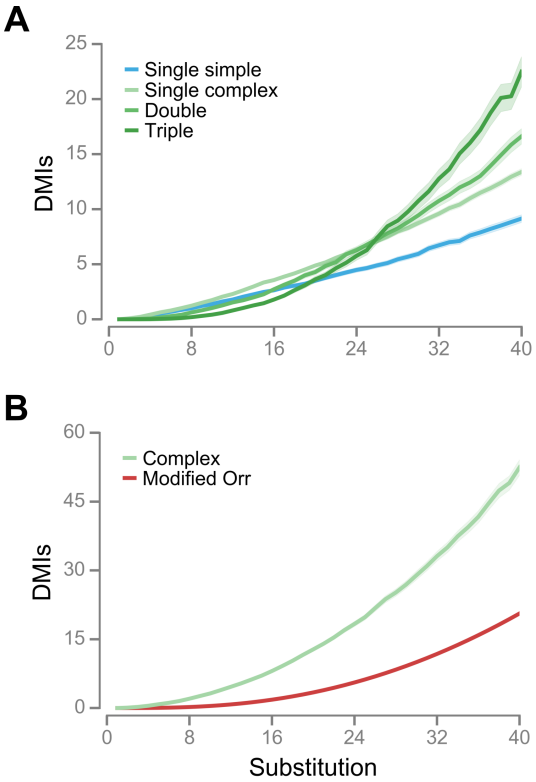

Figure 9 Complex DMIs snowball in the RNA folding model. (A) DMIs inferred through single, double, and triple introgressions. (B) Total number of complex DMIs (green) compared to number predicted if all complex DMIs originate from the "complexification" of simple DMIs and $p = 0.042$ and $q = 0.107$ (red). Values are means of 10^3 stochastic simulations. Shaded regions indicate 95% CIs.

Complex incompatibilities snowball in the RNA folding model

So far we have focused exclusively on simple DMIs. The modified Orr model predicts that complex DMIs should exist if $q > 0$ because they will be generated continuously from simple DMIs. Furthermore, if q is high, the number of complex DMIs should also be high. We tested this prediction in the RNA folding model and found that complex DMIs accumulated in much higher numbers than simple ones: after $k = 40$ substitutions there were approximately 5-fold more complex DMIs than simple ones (Figure 9).

The Orr model predicts that DMIs of order n are expected to accumulate at a rate approximately proportional to k^n (prediction #2: Orr 1995; Welch 2004). Complex DMIs inferred through single introgressions, like simple DMIs, did not show strong support for a snowball effect (Figures 9A and S5; Table S2). However, complex DMIs inferred through double and triple introgressions did snowball, in broad agreement with prediction #2 of the Orr model (Figures 9A and S5; Table S2).

DMIs detected by introgressing i alleles will have order $n \geqslant i + 1$. Thus, prediction #2 of the Orr model leads to the prediction that DMIs inferred by introgressing more alleles are expected to accumulate according to a higher exponent. Our data confirm this prediction qualitatively but not quantitatively: the exponent (c_2) increased with the number of introgressed alleles (i), but $c_2 \ll i + 1$ (Figure 9A; Table S2). Allowing multiple substitutions to occur per site during divergence did not change these results (Figure S6; Table S3). The modified Orr model suggests a possible explanation for these results: complexification affects the accumulation of DMIs of all orders.

Did the complex DMIs in the RNA folding simulations originate from the complexification of simple ones or did they appear *de novo*? If all complex DMIs arise through complexification, then we would expect their number to increase according to the difference between Equations 2 and 9. Figure 9B shows that, although some complex DMIs likely arose from complexification, many complex DMIs must have arisen *de novo*.

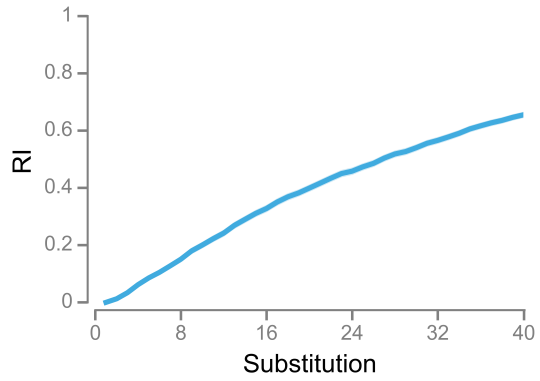

Figure 10 Reproductive isolation (RI) does not snowball in the RNA folding model. Values are means of 10^3 stochastic simulations. Shaded regions indicate 95% CIs.

Reproductive isolation does not snowball in the RNA folding model

Since most DMIs were complex and complex DMIs snowballed, RI would be expected to snowball in the RNA model (prediction #3). However, we found that RI showed a kind of inverse snowball effect—a "slowdown" with divergence. This pattern has been found in many organisms (e.g., Gourbière and Mallet 2010; Giraud and Gourbière 2012). The slowdown was caused by the fact that RI increased slower than linearly with the number of both simple and complex DMIs (Figure S7). Thus, DMIs did not act independently of each other on RI. One likely reason for this non-independence is that the total number of DMIs (simple and complex) among highly diverged sequences is high enough that a substantial fraction of individual sites must participate in multiple DMIs (Figure 9).

The structure of the fitness landscape influences the accumulation of DMIs

Figure 8 shows two striking patterns about the parameters of the modified Orr model. First, p and q were strongly positively correlated with each other (Spearman's rank correlation coefficient: $\rho = 0.466$, $P < 10^{-6}$), indicating that the origination and complexification of simple DMIs are not independent. Second, the parameters varied extensively between simulations. What caused this variation? All simulations took place on the same sequence space, but with different fitness landscapes. Since all fitness landscapes were "holey" (Gavrilets 2004), it follows that the exact pattern of "holeyness" might have had an effect on the evolutionary dynamics. One component of the holeyness of a fitness landscape is the proportion of inviable single mutant neighbors of all the sequences generated during the course of evolution. This measure of the local holeyness of the fitness landscape was strongly positively correlated with both p and q ($\rho = 0.338$ and 0.210, respectively; both, $P < 10^{-6}$) (Figures S8A and S8C).

What determines holeyness? The fitness landscapes in our RNA folding model have two determinants: the reference structure and the value of α (Equation 4). RNA secondary structures can differ in many ways, such as the number and size of base pair stacks, interior loops, and hairpin loops (Schuster *et al.* 1994). The relationship between these structural features and holeyness is difficult to predict *a priori*. For a given reference structure, lower values of α are expected to specify fitness landscapes with more inviable sequences (i.e., holes) in them. To evaluate the extent to which these determinants of the fitness landscape influence holeyness, we ran 10^3 independent evolutionary simulations at each of another four values of α. We found that holeyness was influenced by both determinants of the fitness landscape: it was negatively correlated with α ($\rho = -0.583$; $P < 10^{-6}$; Figure S9A), and positively correlated with the number of base pairs in the reference sequence, β ($\rho = 0.184$; $P < 10^{-6}$; Figure S9B).

The value of α also influenced both the origination and complexification of simple DMIs independently of holeyness. Although α was negatively correlated with holeyness (Figure S9A), which in turn was positively correlated with both p and q (Figures S8A and S8C), α was *positively* correlated with both p and q (Figures S8B and S8D; $\rho = 0.101$ and $\rho = 0.203$ for p and q, respectively; both, $P < 10^{-6}$). The correlations became stronger when we corrected for the effect of holeyness (partial rank correlation coefficients: $\rho = 0.290$ and $\rho = 0.302$ for p and q, respectively; both, $P < 10^{-6}$). Interestingly, changing the value of α had only small effects on the accumulation of simple and complex DMIs (Figure S10).

Discussion

We have tested both predictions and assumptions of the Orr model using a computational model of RNA folding. Our results provide mixed support for the snowball effect (Table 1). Simple DMIs accumulated linearly, contrary to prediction #1 of the Orr model. To elucidate why the snowball appeared to be "missing" from the RNA folding simulations we tested two assumptions of the Orr model. First, that simple DMIs arise with constant probability, p (assumption #1). Although we did detect a decline in p, it was not sufficient to account for the approximately linear pattern of accumulation of simple DMIs. Second, we tested assumption #4 that simple DMIs, once they have arisen, persist indefinitely. We found that this assumption was violated in the RNA folding model. Instead, simple DMIs had a tendency to become more complex as further substitutions took place. We proposed a modified Orr model incorporating the complexification of simple DMIs. The pattern of accumulation of simple DMIs in the RNA folding simulations agrees with this model.

In contrast to simple DMIs, the number of complex DMIs did snowball in the RNA folding simulations, in qualitative agreement with prediction #2 of the Orr model. However, the pattern of accumulation of complex DMIs did not agree with the prediction quantitatively. Despite the snowballing of complex DMIs, RI did not snowball (prediction #3 of the Orr model) because DMIs did not act independently of each other on RI (assumption #3). These results indicate that RI is a poor indicator for the number of DMIs in our model. Thus, the pattern of change in RI with divergence is unsuitable to test the Orr model (Mendelson *et al.* 2004; Johnson 2006; Gourbière and Mallet 2010; Presgraves 2010a). In conclusion, the RNA folding model provided qualitative support for the central prediction of the Orr model that the total number of DMIs snowballs. However, our results failed to confirm certain predictions of the Orr model, as well as some of its assumptions.

An earlier test of the Orr model using a computational model of gene networks also found no evidence for a snowball effect in RI (prediction #3), and concluded that some assumptions of the Orr model were not met (Palmer and Feldman 2009). However, the extent to which the complexification of DMIs influenced their results is unclear because they did not attempt to investigate the dynamics of the DMIs underlying RI.

In one direct empirical test of the snowball effect, DMIs affecting pollen sterility were found to accumulate linearly, whereas DMIs affecting seed sterility were found to accumulate faster than linearly (Moyle and Nakazato 2010). Our results suggest a possible explanation for the discrepancy: faster complexification (i.e., higher q) of pollen sterility DMIs. Sherman *et al.* (2014) found evidence of greater complexity of the DMIs involved in pollen sterility.

If all DMIs are simple and individual loci are at most involved in one DMI, then the proportion of the fixed differences between species where an allele from one species is deleterious in another species, $\mathcal{P}_1$, is expected to increase linearly with genetic distance (Equation 3; Welch 2004). This prediction is contradicted by the observation that $\mathcal{P}_1$ is approximately constant over large genetic distances (Kondrashov *et al.* 2002; Kulathinal *et al.* 2004)—a result we call Welch's paradox. Our results contradict both assumptions behind the prediction that $\mathcal{P}_1$ should increase linearly with genetic distance (Welch 2004): most DMIs are complex, and individual loci are involved in multiple DMIs. These effects are expected to act in opposite directions: the former would cause $\mathcal{P}_1$ to increase faster than linearly with k, whereas the latter would cause $\mathcal{P}_1$ to increase slower than linearly with k. In the RNA folding simulations, $\mathcal{P}_1$ increased with divergence but did so slower than linearly (Figure S11), indicating that the lack of independence between DMIs dominates the evolution of $\mathcal{P}_1$. These results suggest a possible resolution for Welch's paradox: $\mathcal{P}_1$ can be constant even if DMIs snowball if individual loci participate in multiple DMIs. Alternative resolutions of Welch's paradox have been proposed (e.g., Fraïsse *et al.* 2016).

We found that complex DMIs are more abundant than simple DMIs in the RNA folding model. Complex DMIs have been discovered in many introgression studies (reviewed in Wu and Palopoli 1994; Fraïsse *et al.* 2014). For example, Orr and Irving (2001) investigated the sterility of male F1 hybrids between the

Table 1 The RNA folding model provides mixed support for the Orr model

Test	Confirmed? (Data)
Assumption	
1. p constant with divergence	Yes, roughly (Figure 5B)
2. DMIs have small effects on RI	Yes (Figure 10)
3. DMIs have independent effects on RI	No (Figure S7)
4. Simple DMIs persist indefinitely	No (Figures 6, 7B and S2)
Prediction	
1. Simple DMIs snowball	No (Figures 4 and 5A, Table S2)
2. Complex DMIs snowball	Yes, roughly (Figure 9, Table S2)
3. RI snowballs	No (Figure 10)

USA and Bogota subspecies of *D. pseudoobscura* and found that it is caused by an DMI between loci in both chromosomes 2 and 3 of USA and loci in at least three different regions of the X chromosome of Bogota—a DMI of order $n \geqslant 5$. More generally, high-order epistasis appears to be common (Weinreich *et al.* 2013; Kondrashov and Kondrashov 2015; Taylor and Ehrenreich 2015). However, the relative prevalence of simple and complex DMIs in nature is unclear because complex DMIs are more difficult to detect.

Two explanations for the abundance of complex DMIs have been proposed. First, that more complex DMIs evolve more easily than simpler DMIs because they allow a greater proportion of the possible evolutionary paths between the common ancestor and the evolved genotypes containing the DMI (Cabot *et al.* 1994; Orr 1995). Fraïsse *et al.* (2014) tested this mechanism using simulations and concluded that it is unlikely to be effective. Second, that the number of combinations of n loci increases with n (Orr 1995). This explanation is difficult to evaluate in the absence of more information on the probability of origination of complex DMIs. Our results indicate that that probability could be higher than previously thought because complex DMIs are continuously generated from simple DMIs. Indeed, our results suggest a new explanation for the abundance of complex DMIs: that DMIs have a tendency to become increasingly complex with divergence.

Our study has identified one determinant of the origination and complexification of DMIs: the holeyness of the fitness landscape. In a holey fitness landscape, our measure of holeyness is inversely related to the mutational robustness of the genotypes assayed (van Nimwegen *et al.* 1999; Ancel and Fontana 2000). In our model (as in Orr's) "populations" are assumed to contain a single genotype; periodically, a mutant genotype arises and either goes to fixation or disappears. In such a model, mutational robustness is not expected to evolve (van Nimwegen *et al.* 1999). Individual-based simulations would allow us to investigate the intriguing possibility that factors that influence the evolution of mutational robustness (e.g., mutation rate, recombination rate: Wilke *et al.* 2001; Gardner and Kalinka 2006; Azevedo *et al.* 2006) may influence the accumulation of DMIs.

Perhaps the central insight from our study is that simple DMIs have a tendency to become complex. At first glance this claim might seem absurd. Surely a DMI cannot be simple one moment and complex the next. The solution to this puzzle rests, we believe, on the difference between a DMI having a certain order n and our ability to infer that it has order n through genetic crosses. Consider the evolving sequences depicted in Figure 3. Now, imagine that there is a complex DMI of order $n = 3$ between the alleles A_1, B_2, and C_0, and that there are no simple DMIs between pairs of the three alleles (i.e., A_1/B_2, A_1/C_0, and B_2/C_0). For simplicity, we also assume that none of the other alleles at the A, B and C loci are involved in DMIs. The existence of a DMI is defined in the strict sense that any conceivable genotype containing all alleles involved in the DMI is inviable (conversely, the absence of a DMI indicates that at least one of the genotypes containing all alleles involved in the DMI are viable). Despite the $A_1/B_2/C_0$ DMI being complex, after two substitutions ($k = 2$), our introgression and rescue tests would detect a nonexistent simple DMI between alleles A_1 and B_2. Only after the third substitution ($k = 3$) would the true complex DMI be inferred. In the language we have been using so far, the simple DMI would appear to *become* more complex.

The Orr model assumes that it is possible to tell whether a DMI is simple or not. However, a strict definition of "DMI of order n" cannot be applied in practice because the number of genotypes that would have to be tested is astronomically large *and* would have to include mutations that have not even occurred yet. Our protocol for inferring a simple DMI is, as far as we know, the most exhaustive ever devised (the data summarized in Figure 9A required the construction of approximately 6×10^4 introgression genotypes for each individual simulation), but it cannot infer strict-sense simple DMIs. Strict-sense simple DMIs may not even exist in reality. The idea of complexification of DMIs is a natural consequence of using a more practical, broad-sense definition of simple DMI.

We believe that our central finding that simple DMIs have a tendency to become complex is independent of the details of our model. Other results, such as the rate of accumulation of DMIs, are likely to be influenced by the details of our model. The extent to which the RNA folding model is representative of other types of epistatic interactions (e.g., in gene networks) is unclear. One possible criticism is that we used very short sequences and that these are likely to experience unusually strong epistatic interactions. Orr and Turelli (2001) estimated $p \approx 10^{-7}$ in *Drosophila*, a much lower value than found in our simulations. However, an evolution experiment in *S. cerevisiae* detected a simple DMI between two lineages that had only accumulated 6 unique mutations each ($k = 12$) (Anderson *et al.* 2010). This indicates a value of $p \approx 0.015$, within the range of what we observed in the RNA

folding model (Figure 8). Our approach to testing the Orr model can be applied to other computational models of biological systems, such as, transcription-factor binding (Tulchinsky *et al.* 2014; Khatri and Goldstein 2015), gene networks (ten Tusscher and Hogeweg 2009; Palmer and Feldman 2009), and metabolic networks (Barve and Wagner 2013).

Our results were robust to a broad range of holey fitness landscapes defined in the RNA folding model. However, the holey landscape model makes two strong assumptions about the fitness landscape: all viable genotypes had the same fitness, and all low fitness genotypes were completely inviable. Neither assumption is met universally: many alleles involved in DMIs appear to have experienced positive selection during their evolutionary history (Presgraves 2010b; Rieseberg and Blackman 2010; Maheshwari and Barbash 2011), and some DMIs are only mildly deleterious rather than lethal (Presgraves 2003; Schumer *et al.* 2014). These assumptions can be relaxed in the RNA folding model (e.g., Cowperthwaite *et al.* 2005; Draghi *et al.* 2011) and in other models (e.g., Palmer and Feldman 2009; Tulchinsky *et al.* 2014; Khatri and Goldstein 2015).

Studies like ours can test whether the snowball effect occurs under well-defined circumstances. However, they cannot test the *reality* of the snowball effect; introgression studies remain the only way to do so (Matute *et al.* 2010; Moyle and Nakazato 2010; Matute and Gavin-Smyth 2014; Sherman *et al.* 2014; Wang *et al.* 2015).

Acknowledgments

Tim Cooper, Tiago Paixão, Leonie Moyle, and two anonymous reviewers provided useful comments on the manuscript. We had helpful discussions with Rafael Guerrero, Peter Olofsson, and Jeff Tabor. We used the Maxwell and Opuntia clusters from the Center of Advanced Computing and Data Systems (CACDS) at the University of Houston. CACDS staff provided technical support. The National Science Foundation (grant DEB-1354952 awarded to R.B.R.A.) funded this work.

Literature Cited

Ancel, L. W. and W. Fontana, 2000 Plasticity, evolvability, and modularity in RNA. J. Exp. Zool. (Mol. Dev. Evol.) **288**: 242–283.

Anderson, J. B., J. Funt, D. A. Thompson, S. Prabhu, A. Socha, C. Sirjusingh, J. R. Dettman, L. Parreiras, D. S. Guttman, A. Regev, and L. M. Kohn, 2010 Determinants of divergent adaptation and Dobzhansky-Muller interaction in experimental yeast populations. Curr. Biol. **20**: 1383 – 1388.

Azevedo, R. B. R., R. Lohaus, S. Srinivasan, K. K. Dang, and C. L. Burch, 2006 Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. Nature **440**: 87–90.

Barve, A. and A. Wagner, 2013 A latent capacity for evolutionary innovation through exaptation in metabolic systems. Nature **500**: 203–206.

Bolnick, D. I. and T. J. Near, 2005 Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). Evolution **59**: 1754–1767.

Cabot, E. L., A. W. Davis, N. A. Johnson, and C. I. Wu, 1994 Genetics of reproductive isolation in the *Drosophila simulans* clade: complex epistasis underlying hybrid male sterility. Genetics **137**: 175–189.

Cowperthwaite, M. C., J. J. Bull, and L. A. Meyers, 2005 Distributions of beneficial fitness effects in RNA. Genetics **170**: 1449–57.

Coyne, J. A. and H. A. Orr, 1989 Patterns of speciation in *Drosophila*. Evolution **43**: 362–381.

Darwin, C., 1859 *On the Origin of Species by Means of Natural Selection*. J. Murray, London,.

Dettman, J. R., D. J. Jacobson, E. Turner, A. Pringle, and J. W. Taylor, 2003 Reproductive isolation and phylogenetic divergence in *Neurospora*: Comparing methods of species recognition in a model eukaryote. Evolution **57**: 2721–2741.

Dobzhansky, T., 1937 *Genetics and the Origin of Species*. Columbia Univ. Press, New York.

Doshi, K. J., J. J. Cannone, C. W. Cobaugh, and R. R. Gutell, 2004 Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. BMC Bioinformatics **5**: 105.

Draghi, J. A., T. L. Parsons, and J. B. Plotkin, 2011 Epistasis increases the rate of conditionally neutral substitution in an adapting population. Genetics **187**: 1139–52.

Draghi, J. A., T. L. Parsons, G. P. Wagner, and J. B. Plotkin, 2010 Mutational robustness can facilitate adaptation. Nature **463**: 353–355.

Edmands, S., 2002 Does parental divergence predict reproductive compatibility? Tr. Ecol. Evol. **17**: 520–527.

Fraïsse, C., J. A. D. Elderfield, and J. J. Welch, 2014 The genetics of speciation: Are complex incompatibilities easier to evolve? J. Evol. Biol. **27**: 688–699.

Fraïsse, C., P. A. Gunnarsson, D. Roze, N. Bierne, and J. J. Welch, 2016 The genetics of speciation: Insights from Fisher's geometric model. Evolution **70**: 1450–1464.

Gardner, A. and A. T. Kalinka, 2006 Recombination and the evolution of mutational robustness. J. Theor. Biol. **241**: 707–715.

Gavrilets, S., 2004 *Fitness Landscapes and the Origin of Species*. Princeton Univ. Press.

Giraud, T. and S. Gourbière, 2012 The tempo and modes of evolution of reproductive isolation in fungi. Heredity **109**: 204–214.

Gourbière, S. and J. Mallet, 2010 Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball". Evolution **64**: 1–24.

Jewell, C., A. D. Papineau, R. Freyre, and L. C. Moyle, 2012 Patterns of reproductive isolation in *Nolana* (Chilean bellflower). Evolution **66**: 2628–2636.

Johnson, N. A., 2006 The evolution of reproductive isolating barriers. In *Evolutionary Genetics: Concepts and Case Studies*, edited by C. W. Fox and J. B. Wolf, pp. 374–398, Oxford Univ. Press, Oxford, U.K.

Kachroo, A. H., J. M. Laurent, C. M. Yellman, A. G. Meyer, C. O. Wilke, and E. M. Marcotte, 2015 Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science **348**: 921–925.

Khatri, B. S. and R. A. Goldstein, 2015 Simple biophysical model predicts faster accumulation of hybrid incompatibilities in small populations under stabilizing selection. Genetics **201**: 1525–37.

Kim, S., 2015 ppcor: An R package for a fast calculation to semi-partial correlation coefficients. Comm. Stat. Appl. Meth. **22**: 665–674.

Kondrashov, A. S., S. Sunyaev, and F. A. Kondrashov, 2002 Dobzhansky-Muller incompatibilities in protein evolution.

Proc. Natl. Acad. Sci. U. S. A. **99**: 14878–14883.
Kondrashov, D. A. and F. A. Kondrashov, 2015 Topological features of rugged fitness landscapes in sequence space. Tr. Genet. **31**: 24–33.
Kulathinal, R. J., B. R. Bettencourt, and D. L. Hartl, 2004 Compensated deleterious mutations in insect genomes. Science **306**: 1553–1554.
Larcombe, M. J., B. Holland, D. a. Steane, R. C. Jones, D. Nicolle, R. E. Vaillancourt, and B. M. Potts, 2015 Patterns of reproductive isolation in *Eucalyptus*—a phylogenetic perspective. Molecular Biology and Evolution **32**: 1833–1846.
Lijtmaer, D. A., B. Mahler, and P. L. Tubaro, 2003 Hybridization and postzygotic isolation patterns in pigeons and doves. Evolution **57**: 1411–1418.
Liti, G., D. B. H. Barton, and E. J. Louis, 2006 Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. Genetics **174**: 839–850.
Livingstone, K., P. Olofsson, G. Cochran, A. Dagilis, K. MacPherson, and K. A. Seitz Jr., 2012 A stochastic model for the development of Bateson–Dobzhansky–Muller incompatibilities that incorporates protein interaction networks. Math. Biosci. **238**: 49 – 53.
Lorenz, R., S. H. Bernhart, C. Höner Zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker, 2011 ViennaRNA Package 2.0. Algorithms Mol. Biol. **6**: 26.
Maheshwari, S. and D. A. Barbash, 2011 The genetics of hybrid incompatibilities. Annu. Rev. Genet. **45**: 331–355.
Mathews, D. H., J. Sabina, M. Zuker, and D. H. Turner, 1999 Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. **288**: 911–940.
Matute, D. R., I. A. Butler, D. A. Turissini, and J. A. Coyne, 2010 A test of the snowball theory for the rate of evolution of hybrid incompatibilities. Science **1518**.
Matute, D. R. and J. Gavin-Smyth, 2014 Fine mapping of dominant X-linked incompatibility alleles in *Drosophila* hybrids. PLoS Genet. **10**.
Mendelson, T. C., 2003 Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae: *Etheostoma*). Evolution **57**: 317–327.
Mendelson, T. C., B. D. Inouye, and M. D. Rausher, 2004 Quantifying patterns in the evolution of reproductive isolation. Evolution **58**: 1424–1433.
Moyle, L. C. and T. Nakazato, 2010 Hybrid incompatibility "snowballs" between *Solanum* species. Science **329**: 1521–1523.
Moyle, L. C., M. S. Olson, and P. Tiffin, 2004 Patterns of reproductive isolation in three angiosperm genera. Evolution **58**: 1195–1208.
Muller, H. J., 1942 Isolating mechanisms, evolution and temperature. Biol. Symp. **6**: 71–125.
Nosil, P. and J. L. Feder, 2012 Genomic divergence during speciation: causes and consequences. Phil. Trans. R. Soc. B **367**: 332–342.
Orr, H. A., 1995 The population genetics of speciation: The evolution of hybrid incompatibilities. Genetics **139**: 1805–1813.
Orr, H. A. and S. Irving, 2001 Complex epistasis and the genetic basis of hybrid sterility in the *Drosophila pseudoobscura* Bogota-USA hybridization. Genetics **158**: 1089–1100.
Orr, H. A. and M. Turelli, 2001 The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. Evolution **55**: 1085–1094.
Palmer, M. E. and M. W. Feldman, 2009 Dynamics of hybrid incompatibility in gene networks in a constant environment. Evolution **63**: 418–431.
Presgraves, D. C., 2002 Patterns of postzygotic isolation in Lepidoptera. Evolution **56**: 1168–1183.
Presgraves, D. C., 2003 A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. Genetics **163**: 955–972.
Presgraves, D. C., 2010a Speciation genetics: search for the missing snowball. Curr. Biol. **20**: R1073–4.
Presgraves, D. C., 2010b The molecular evolutionary basis of species formation. Nat. Rev. Gen. **11**: 175–180.
R Core Team, 2016 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
Rieseberg, L. H. and B. K. Blackman, 2010 Speciation genes in plants. Ann. Bot. **106**: 439–455.
Roberts, M. S. and F. M. Cohan, 1993 The effect of DNA sequence divergence on sexual isolation in *Bacillus*. Genetics **134**: 401–408.
Sasa, M. M., P. T. Chippindale, and N. A. Johnson, 1998 Patterns of postzygotic isolation in frogs. Evolution **52**: 1811–1820.
Schumer, M., R. Cui, D. Powell, R. Dresner, G. G. Rosenthal, and P. Andolfatto, 2014 High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. eLife .
Schuster, P., W. Fontana, P. F. Stadler, and I. L. Hofacker, 1994 From sequences to shapes and back: a case study in RNA secondary structures. Proc. R. Soc. B **255**: 279–284.
Scopece, G., A. Musacchio, A. Widmer, and S. Cozzolino, 2007 Patterns of reproductive isolation in mediterranean deceptive orchids. Evolution **61**: 2623–2642.
Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Saetre, C. Bank, A. Brännström, A. Brelsford, C. S. Clarkson, F. Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, P. Franchini, C. D. Jiggins, F. C. Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. A. Marques, S. H. Martin, B. Matthews, J. I. Meier, M. Möst, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, E. T. Watson, A. M. Westram, and A. Widmer, 2014 Genomics and the origin of species. Nat. Rev. Genet. **15**: 176–92.
Sherman, N. A., A. Victorine, R. J. Wang, and L. C. Moyle, 2014 Interspecific tests of allelism reveal the evolutionary timing and pattern of accumulation of reproductive isolation mutations. PLoS Genet **10**: e1004623.
Stelkens, R. B., K. A. Young, and O. Seehausen, 2010 The accumulation of reproductive incompatibilities in african cichlid fish. Evolution **64**: 617–633.
Taylor, M. B. and I. M. Ehrenreich, 2015 Higher-order genetic interactions and their contribution to complex traits. Tr. Genet. **31**: 34–40.
ten Tusscher, K. H. W. J. and P. Hogeweg, 2009 The role of genome and gene regulatory network canalization in the evolution of multi-trait polymorphisms and sympatric speciation. BMC Evol. Biol. **9**: 159.
The Marie Curie SPECIATION Network, 2012 What do we need to know about speciation? Tr. Ecol. Evol. **27**: 27–39.
Tulchinsky, A. Y., N. A. Johnson, W. B. Watt, and A. H. Porter, 2014 Hybrid incompatibility arises in a sequence-based bioenergetic model of transcription factor binding. Genetics **198**: 1155–1166.
van Nimwegen, E., J. P. Crutchfield, and M. Huynen, 1999 Neutral evolution of mutational robustness. Proc. Natl. Acad. Sci. U. S. A. **96**: 9716–9720.

Wagner, A., 2008 Robustness and evolvability: a paradox resolved. Proc. R. Soc. B **275**: 91–100.

Wang, R. J., M. A. White, and B. A. Payseur, 2015 The pace of hybrid incompatibility evolution in house mice. Genetics **201**: 229–242.

Weinreich, D. M., Y. Lan, C. S. Wylie, and R. B. Heckendorn, 2013 Should evolutionary geneticists worry about higher-order epistasis? Curr. Opin. Genet. Dev. **23**: 700–707.

Welch, J. J., 2004 Accumulating Dobzhansky-Muller incompatibilities: Reconciling theory and data. Evolution **58**: 1145–1156.

Wilke, C. O., J. L. Wang, C. Ofria, R. E. Lenski, and C. Adami, 2001 Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature **412**: 331–333.

Wu, C.-I. and M. F. Palopoli, 1994 Genetics of postmating reproductive isolation in animals. Annu. Rev. Genet. **28**: 283–308.