bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The Melting Snowball: A Test of the Snowball Model Using RNA

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, alleles at different loci that have fixed in different genetic backgrounds may be incompatible when brought together in a hybrid. 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 model using data from natural populations. One limitation of these studies is that they have focused on predictions of the snowball model but not on its underlying assumptions. Here we use a computational model of RNA folding to test both predictions and assumptions of the snowball model. In this 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 snowball model. However, simple DMIs can disappear after they have arisen, contrary to the assumptions of the snowball 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 snowball model—the melting snowball model—where simple DMIs can become complex after they appear. The melting snowball 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 snowball model that the number of DMIs snowballs, but challenges some of its underlying assumptions.

KEYWORDS speciation; Dobzhansky-Muller incompatibility; intrinsic postzygotic isolation; snowball effect; RNA folding; 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 reproducitve 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, 1997; Sasa *et al.* 1998; Edmands 2002; Fitzpatrick 2002; Presgraves 2002; Lijtmaer *et al.* 2003; Mendelson *et al.* 2004; Bolnick and Near 2005; Johnson 2006; Gourbière and Mallet 2010; Giraud and Gourbière 2012). 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 mechanism through which genetic divergence can cause RI.

Copyright © 2016 by the Genetics Society of America
doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Monday 19th September, 2016%
[1]Corresponding author: Department of Biology and Biochemistry, University of Houston, Houston, TX, US. E-mail: razevedo@uh.edu

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

They noted that, in the absence of gene flow between two populations, new mutations that fix at different loci may interact negatively with one another when 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 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, a pattern Orr (1995) described as "snowballing." This prediction assumes that p remains constant as populations diverge.

DMIs involving 3 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 (Orr 1995; Welch 2004).

Several studies have attempted to test the snowball model. They have employed three different approaches. First, using postzygotic RI as a proxy for the number of DMIs. 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 the prediction of the snowball model. Results from other studies using a similar approach have provided little support for the snowball model (Coyne and Orr 1989, 1997; Sasa *et al.* 1998; Fitzpatrick 2002; Presgraves 2002; Lijtmaer *et al.* 2003; Mendelson *et al.* 2004; Bolnick and Near 2005; Gourbière and Mallet 2010; Giraud and Gourbière 2012), leading some to pronounce the snowball "missing" (Johnson 2006; Gourbière and Mallet 2010). However, this indirect approach cannot provide a strong test of the snowball model because it relies on the untested ancillary assumption of a linear relationship between the number of DMIs and RI. This assumption will not be met if, for example, DMIs do not act independently on RI (Orr 1995; Welch 2004).

The second approach to testing the snowball model involves estimating the number of DMIs directly. For example, Moyle and Nakazato (2010) used a QTL mapping approach to test the snowball model in species of *Solanum*. 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, in agreement with the prediction of the snowball model. However, DMIs affecting pollen sterility appeared to accumulate linearly, contrary to the snowball model. Studies following this second approach (Matute *et al.* 2010; Moyle and Nakazato 2010; Matute and Gavin-Smyth 2014; Sherman *et al.* 2014; Wang *et al.* 2015) 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 and final approach is ingenious because it does not require the direct study of hybrids. 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. If each allele participates at most in one DMI, then we have $\mathcal{P}_1 = I_k/k$. This relationship assumes that p is low. 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 snowball model (Equation 2; Welch 2004)

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

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 the prediction of the snowball model (Welch 2004; Fraïsse *et al.* 2016).

The tests of the snowball model outlined above give inconsistent results. Specifically, the most direct approaches (i.e., the second and third) give opposite results, a paradox first noted by Welch (2004). One common limitation to all approaches is that they focus on testing predictions of the snowball model, without testing its assumptions (e.g., constant p). Furthermore, each approach makes additional assumptions that also go untested (e.g., DMIs act independently on RI). Here we use a computational model of RNA folding (Schuster *et al.* 1994) to test both predictions and assumptions of the snowball model. The RNA folding model 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). Our results provide mixed support for the snowball 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.

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

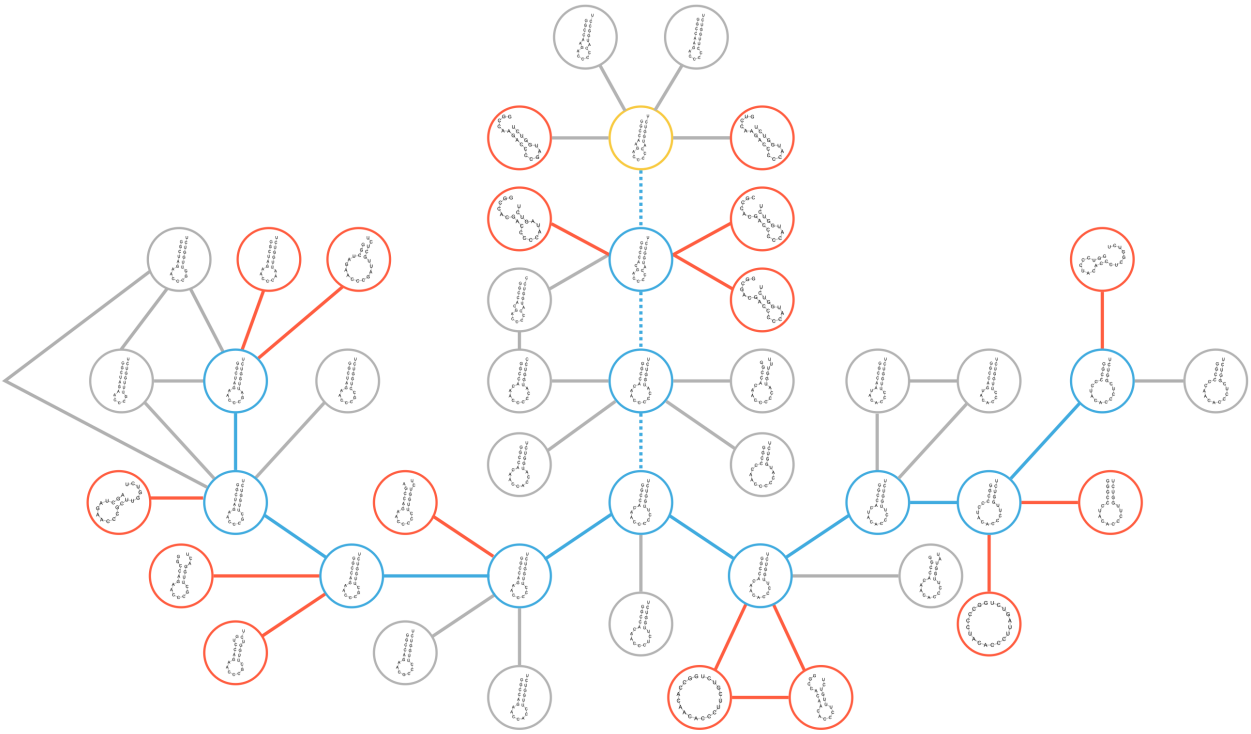

Figure 1 Evolution on a holey fitness landscape. Mutational network of RNA sequences. Lines connect sequences of 20 nucleotides that can be reached by a single nucleotide substitution. Only a tiny fraction of the entire mutational network of $\sim 10^{12}$ sequences is shown. Furthermore, only a few 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 (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 (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 (solid blue lines).

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (Figure 1; Gavrilets 2004).

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 Orr's (1995) model. All types of mutations have equal probability.

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.

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

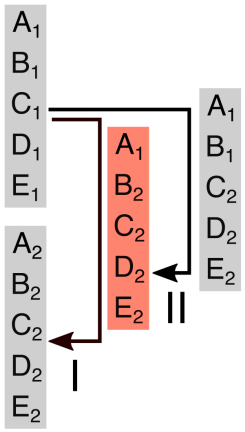

Figure 2 Detecting DMIs. To find simple DMIs, we use an introgression–rescue assay where we introgress one diverged allele between the two lineages (step I: $1 \xrightarrow{A} 2$), and if this substitution results in an inviable genotype (red), we try to rescue it with a second introgression (step II: $1 \xrightarrow{A,B} 2$). 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.

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 its expression depends on the genetic background.

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 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 $k \geqslant 2$.

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).

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.

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 to 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 the 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 two 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.

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Direct simulation of the snowball model

We also simulate the accumulation of DMIs following the snowball model (Orr 1995). 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 snowball 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 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). Semi-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 will be made available at the time of publication at https://github.com/. The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

Results

Simple DMIs do not snowball in the RNA folding model

The snowball 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 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

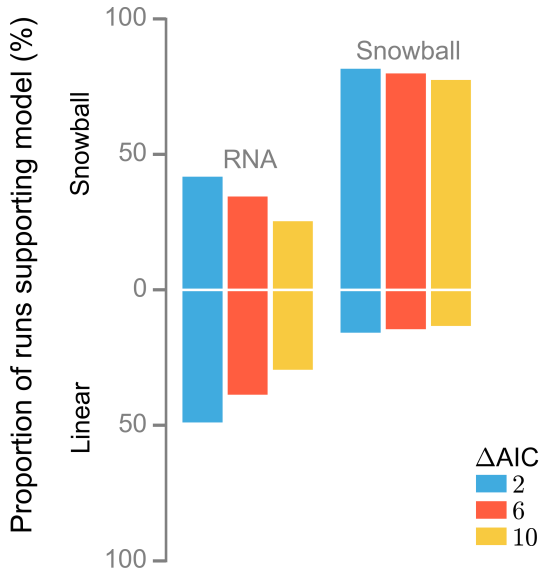

Figure 4 Simple DMIs do not snowball in the RNA folding model. We fitted the snowball model (Equation 2) and a linear model (Equation 6) to each run from two kinds of simulations: simulations of the RNA folding model, and direct simulations of the snowball process with values of p estimated by fitting the model in Equation 2 to each RNA folding simulation (Figure S1A). Bars above (below) the x-axis show the proportions of runs providing stronger support for the snowball (linear) model; the proportions of runs providing approximately equal support for both models are not shown. Each proportion is based on 10^3 stochastic simulations. Different colors indicate the ΔAIC thresholds used in evaluating the level of support for the two models.

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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). The average response in the number of DMIs in the RNA folding simulations was approximately linear (Figure 5A), in agreement with the AIC analysis.

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 the snowball process over $k = 40$ substitutions assuming values of p estimated by fitting the snowball model in Equation 2 to the RNA folding data (Figure S1A). As expected, these direct snowball simulations provided much stronger support for the snowball model than the RNA simulations (Figure 4). We conclude that simple DMIs do not snowball in at least some RNA folding simulations.

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

What explains the lack of support for the snowball model in the RNA folding simulations? One possibility is that p itself evolved, contrary to the assumption of the snowball model (Orr 1995).

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 snowball model may be violated in the RNA folding model: that simple DMIs, once they have arisen, persist indefinitely. This assumption is implicit in the original description of the snowball model (Orr 1995) and, to our knowledge, has never been called into question.

To test this assumption, 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.

The snowball model assumes that the DMI network is static: as populations evolve they actualize potential DMIs (for an

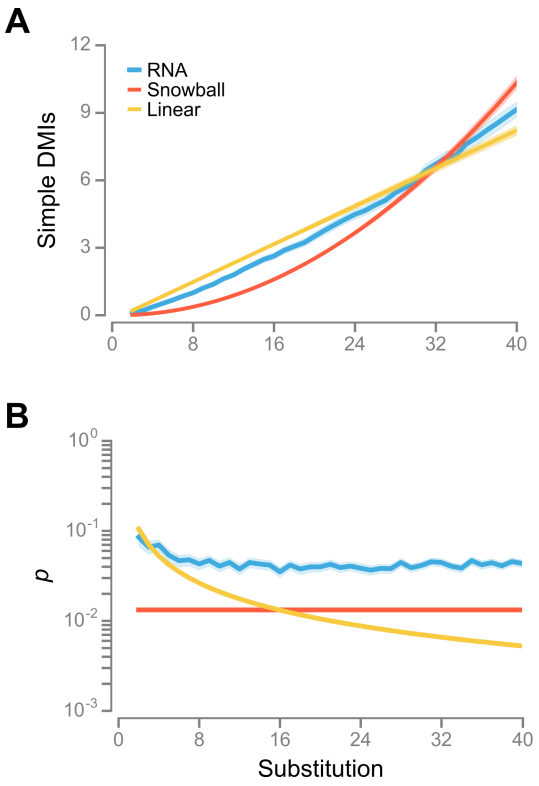

Figure 5 Simple DMIs do not snowball in the RNA folding model. (A) Evolution of the number of simple DMIs, I_k, as two populations diverge by accumulating substitutions, k. Values are means of 10^3 runs of three different kinds of stochastic simulations: "RNA," simulations of the RNA folding model (blue); "snowball," direct simulations of the snowball process with constant p estimated as explained in (B) (red); "linear," direct simulations of the snowball process with declining p 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.

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

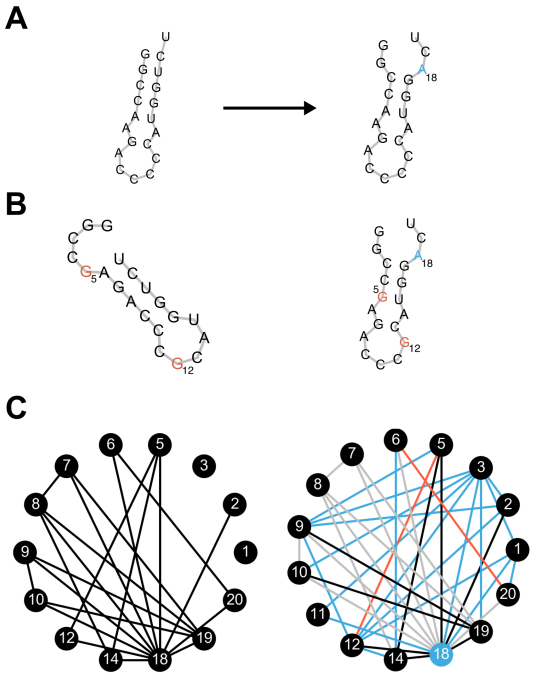

Figure 6 A single substitution can dramatically rearrange the network of potential DMIs. (A) The 20 nucleotide long RNA sequence on the left acquires a neutral U→A substitution at position 18 (blue). The holey fitness landscape is defined by $\alpha = 2$ (Equation 4). The secondary structure of the sequence 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 losses of potential simple DMIs in the network on the right. Gray edges indicate losses due to the constituent alleles no longer being neutral in the new background. Red edges indicate losses caused by complexification; the DMI discussed in (B) is an example (5–12 edge). The Jaccard index (Equation 5) between the two networks is $J = 0.205$.

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 "loss" 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 loss 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 losses do not imply that DMIs cannot persist indefinitely. However, if there is a bias towards such losses of potential DMIs relative to gains of the same kind then p is expected to decline with divergence. The majority of losses in Figure 6C (gray lines) are of this kind.

The second kind of loss 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 is lost in this way. This kind of loss—complexification—implies that some simple DMIs may not persist indefinitely.

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 snowballing.

The melting snowball model

We incorporate the dynamic nature of simple DMIs by extending the snowball 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 metaphor evokes a snowball rolling down a hillside, picking up snow (appearance of simple DMIs) as it rolls, causing it to increase in size. To stretch the metaphor, we call the new model the *melting snowball*: as the snowball rolls it also melts (complexification of simple DMIs), causing it to decrease in size. Neither metaphor should be taken too literally, though. For example, both metaphors give the mistaken impression that the accumulation of DMIs itself *causes* the emergence of new DMIs, which is not part of either model.

The snowball model is a special case of the melting snowball 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 (Figure 7A). First, the increase in the number of simple DMIs eventually becomes linear

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

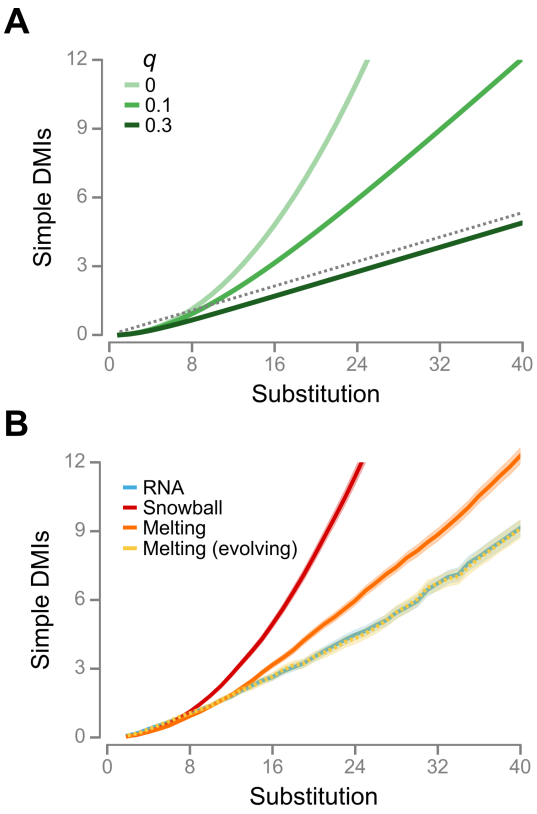

Figure 7 The RNA folding simulations agree with the melting snowball model. (A) Evolution of the number of simple DMIs under the melting snowball model. Responses for $p = 0.04$ and different values of q. The dashed line shows a slope of p/q for $q = 0.3$. (B) Mean responses of 10^3 runs of four different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

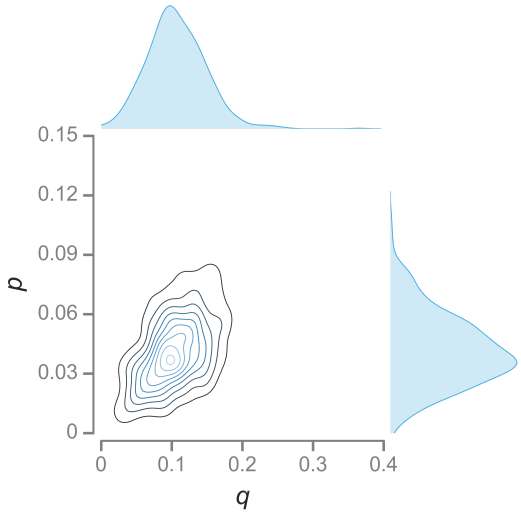

Figure 8 Distributions of the parameters of the melting snowball model in the RNA folding simulations: p, the probability that a simple DMI arises, and q, the probability that a simple DMI becomes complex. One- and two-dimensional kernel density estimates based on 10^3 stochastic simulations. For each 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 melting snowball 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 melting snowball 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 linear response in the number of simple DMIs in the RNA folding simulations can be explained by the melting snowball model.

To evaluate the extent to which the melting snowball model can account for the lack of support for the snowball model in our RNA folding simulations, we conducted 10^3 direct simulations of the melting snowball process 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 melting snowball simulations 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 melting snowball model explains the RNA folding results.

Figure 7B shows that the melting snowball model (orange) approximates the RNA folding data better than the snowball 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 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 melting snowball process matched the RNA folding data perfectly (Figure 7B). We conclude that the melting snowball model explains the results of the RNA folding model, provided we relax the assumptions that p and q are constant.

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

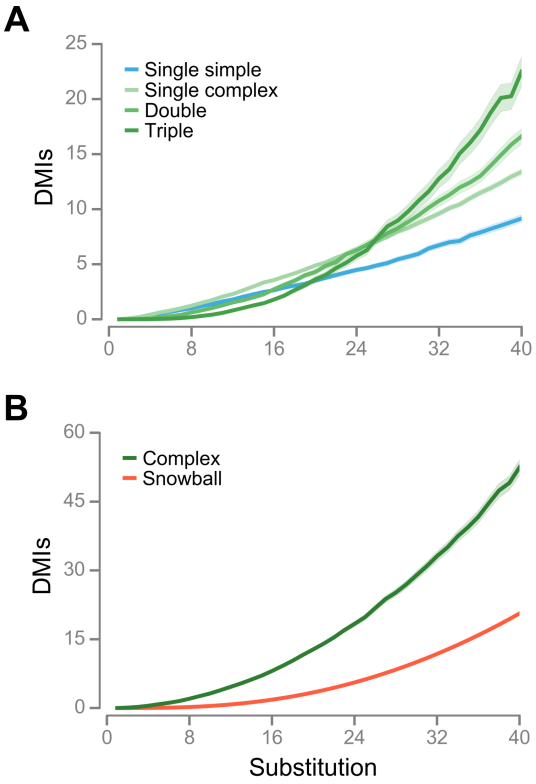

Figure 9 Complex DMIs snowball in the RNA folding model. (A) DMIs inferred through single, double, and triple introgressions. (B) Total number of complex DMIs (green) compared to number predicted if all complex DMIs originate from the "melting" of simple DMIs and $p = 0.042$ and $q = 0.107$ (red). Values 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 melting snowball 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 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 snowball model predicts that the number of complex DMIs should snowball (Orr 1995; Welch 2004). Complex DMIs, unlike simple ones, did snowball (Figures 9 and S5). In addition, complex DMIs detected by introgressing more alleles accumulated faster (Figure 9B). Allowing multiple substitutions to occur per site during divergence did not change this pattern (Figure S6). These results indicate that higher-order DMIs accumulated faster than lower-order DMIs.

Did the complex DMIs originate from the "melting" of simple ones or did they appear *de novo*? If all complex DMIs arise through melting, 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 melting, many complex DMIs must have arisen *de novo*.

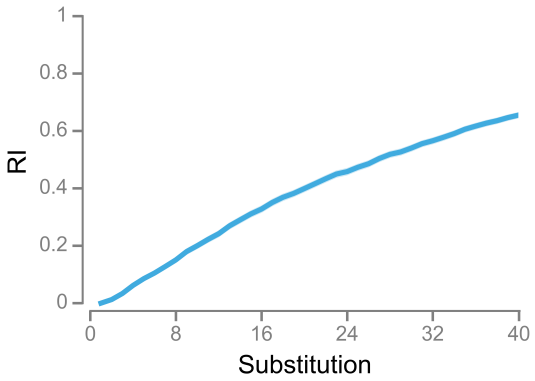

Figure 10 Reproductive isolation (RI) does not snowball in the RNA folding model. Values are means of 10^3 stochastic simulations. Shaded regions indicate 95% CIs.

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. However, we found that RI showed a kind of inverse snowball—a "slowdown" with divergence. This pattern has been found in many organisms (e.g., Gourbière and Mallet 2010; Giraud and Gourbière 2012). This 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 fitness landscape influences the parameters of the melting snowball model

Figure 8 shows two striking patterns about the parameters of the melting snowball model. First, p and q were strongly positively

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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? 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). 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 test whether 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 (Figure S9): it was positively correlated with the number of base pairs in the reference sequence ($\rho = 0.184$; $P < 10^{-6}$) and negatively correlated with α ($\rho = -0.583$; $P < 10^{-6}$).

Changing α did not affect the patterns of accumulation of simple and complex DMIs qualitatively (Figure S10). Interestingly, α was strongly positively correlated with both p and q (Figures S8B and S8D): the semi-partial rank correlation coefficient when the effect of holeyness was removed from α were $\rho = 0.282$ for p and $\rho = 0.301$ for q (both, $P < 10^{-6}$). This result is counterintuitive because α was negatively correlated with holeyness, which in turn was positively correlated with both p and q. We conclude that the parameters of the melting snowball model were influenced independently by both holeyness and α.

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

Test	Confirmed?
Prediction	
Simple DMIs snowball	No
Complex DMIs snowball	Yes
RI snowballs	No
Assumption	
Constant p with divergence	Yes, roughly
Simple DMIs persist indefinitely	No
Linear relationship between number of DMIs and RI	No

Discussion

We have tested both predictions and assumptions of the snowball model using a computational model of RNA folding. Our results provide mixed support for the snowball model (Table 1). Simple DMIs accumulated linearly, contrary to one of the main quantitative predictions of the snowball model (Orr 1995) (Figures 4 and 5A). To elucidate why the snowball appeared to be "missing" from the RNA folding simulations we tested two assumptions of the snowball model. First, that simple DMIs arise with constant probability, p. Although we did detect a decline in p (Figure 5B), it was not sufficient to account for the approximately linear pattern of accumulation of simple DMIs. Second, we tested the assumption 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 conclude that the snowball was "melting" for simple DMIs, not missing.

We proposed an extended snowball model incorporating the complexification of simple DMIs—the melting snowball. The RNA folding simulations agree with this model. In contrast to simple DMIs, the number of complex DMIs did snowball, in agreement with the prediction of the snowball model. In conclusion, the RNA folding model supported the central prediction of the snowball model that the number of DMIs snowballs, but challenged some of its underlying assumptions.

Despite the snowballing of DMIs, RI did not snowball because DMIs did not act independently of each other on RI. 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 snowball model (Johnson 2006; Gourbière and Mallet 2010; Presgraves 2010a).

In one direct test of the snowball model, DMIs affecting pollen sterility were found to accumulate linearly (Moyle and Nakazato 2010). However, the extent to which those results might be explained by the melting snowball model is unclear for two reasons. First, the order of the DMIs detected in that study is unknown. Second, DMIs may have been missed if individual alleles participated in multiple DMIs. Our study avoided these problems for simple DMIs. A linear accumulation of simple DMIs can occur in the presence of gene flow (Kondrashov 2003), which was not incorporated in our model.

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 (Welch 2004). Our results contradict both assumptions behind the prediction that $\mathcal{P}_1$ should increase linearly with genetic distance: 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

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 Fraïsse *et al.* 2014). For example, Orr and Irving (2001) investigated the sterility of male F1 hybrids between the 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.

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 snowball model (Orr 1995) 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 simple DMIs in the strict sense. Simple DMIs in the strict sense may not even exist. The idea of complexification of DMIs is a natural consequence of using a more practical, broad-sense definition of simple DMI.

The extent to which the RNA folding model is representative of other types of epistatic interactions 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 *Saccharomyces 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 $p = 0.015$, within the range of what we observed in the RNA folding model (Figure 8).

We found that 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). Other fitness landscapes can be implemented readily within the RNA folding model (e.g., Cowperthwaite *et al.* 2005; Draghi *et al.* 2011). The extent to which relaxing the assumptions of the holey landscape model will affect our results is a question for future research.

Acknowledgments

Tim Cooper and Tiago Paixão gave useful comments on the manuscript. We had helpful discussions with Rafael Guerrero and Peter Olofsson. 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.

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.

Coyne, J. A. and H. A. Orr, 1997 "Patterns of speciation in *Drosophila*" revisited. Evolution **51**: 295–303.

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

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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.

Fitzpatrick, B. M., 2002 Molecular correlates of reproductive isolation. Evolution **56**: 191–198.

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.

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.

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.

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., 2003 Accumulation of Dobzhansky-Muller incompatibilities within a spatially structured population. Evolution **57**: 151–153.

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.

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.

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., 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.

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.

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.

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.

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.

Taylor, M. B. and I. M. Ehrenreich, 2015 Higher-order genetic interactions and their contribution to complex traits. Tr. Genet. **31**: 34–40.

bioRxiv preprint doi: https://doi.org/10.1101/076232; this version posted September 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The Marie Curie SPECIATION Network, 2012 What do we need to know about speciation? Tr. Ecol. Evol. **27**: 27–39.

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.