Pairwise hybrid incompatibilities dominate during allopatric speciation for a simple genotype-phenotype map of embryonic spatial patterning

Bhavin S. Khatri[*†,1] **and Richard A. Goldstein**[†]

[*]The Francis Crick Institute, London, United Kingdom, [†]Division of Infection & Immunity, University College London, London, United Kingdom

ABSTRACT Understanding the origin of species is as Darwin called it "that mystery of mysteries". Yet still, how the processes of evolution give rise to non-interbreeding species is not well understood. In an empirical search for a genetic basis, transcription factor DNA binding has commonly been identified as being an important factor in the development of reproductive isolation. This is supported by computational and theoretical models based on the biophysics of transcription factor DNA binding that provide a mechanistic basis of such incompatibilities between allopatrically evolving populations. However, gene transcription mediated by such binding events occurs within the context of larger gene regulatory networks, so the question remains how important are such pair-wise interactions compared to higher order interactions in determining incompatibilities. Orr calculated that as the order of interaction increases there are more pathways for an incompatibility to occur. Here, we show, using simulations based on a simple biophysical genotype phenotype map of spatial patterning in development, that biophysics provides a stronger constraint, leading to pair-wise incompatibilities arising more quickly and being more numerous than higher order incompatibilities, when there is stabilising selection on each allopatric lineage. In addition, we show that incompatibilities arise more quickly for smaller populations and in a manner supporting previous conclusions from models of hybrid incompatibility based solely on transcription factor DNA binding.

KEYWORDS speciation, Dobzhansky-Muller incompatibilities, sequence entropy, population size, co-evolution, genotype phenotype map

Introduction

The detailed genetic mechanisms by which non-interbreeding species arise is still largely not understood. Darwin called it that "mystery of mysteries" Darwin (1859); he struggled to understand how natural selection could give rise to hybrid inviability or infertility. In a modern setting, Darwin's conundrum was, if a hybrid incompatibility were due to a single locus, how could two species, fixed for AA and aa, respectively, evolve from a common ancestor, since at least one of these species would need the inviable genotype Aa to give rise to offspring. A solution to this problem was conceived independently by Dobzhansky, Muller and Bateson Dobzhansky (1936); Muller (1942); Bateson (1909), by which neither population need pass through a bottleneck. If instead incompatibilities arise due to non-linear or epistatic fitness interactions between different loci, it is possible for example, that two lines that are geographically isolated from each other, evolve independently from a common ancestor ab (allopatric evolution), fix the allelic combinations aB and Ab respectively, yet the hybrid genotype AB is inviable.

The work of Orr provided a framework to understand how incompatibilities might arise in allopatry Orr (1995); Orr and Turelli (2001), when populations are small, $\mu N \ll 1$, (where μ is the mutation rate for a typical loci and N the effective population size) and essentially monomorphic. Orr suggested that as two lines fix independent substitutions from a common ancestor, any combination of alleles that may arise in hybrids that have not been 'tested' or explicitly fixed by the process of evolution, could potentially cause an incompatibility. Assuming an

doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Sunday 2nd April, 2017%
[1]Division of Infection & Immunity, University College London & The Francis Crick Institute, London, United Kingdom, bhavin.khatri@physics.org

infinite number of loci and that back-substitutions or multiple substitutions at the same loci are not possible, Orr showed that the number of incompatibilities involving n-loci increases as $\sim K^n$, for K substitutions separating the two lines and assuming $n \ll K$. In particular, pair-wise incompatibilities ($n = 2$) rise as K^2. This has been likened to a "snowball" effect as the number of incompatibilities rises rapidly with the number of substitutions that separate the lines. As an outcome of this framework Orr also hypothesised that more complex DMIs, which involve more than 2-loci should be easier to evolve as the fraction viable paths between common ancestor genotype and the genotypes of the two present day species is larger Orr (1995) for a fixed number of incompatible genotypes. It is, however, an open question, whether these predictions would hold for more realistic fitness landscapes.

To address the question of how incompatibilities develop in more realistic fitness landscapes, Tulchinsky et al Tulchinsky *et al.* (2014b,a) developed sequence-based simulations that investigated the mechanistic basis of the evolution of hybrid incompatibilities and the effects of pleiotropy. Khatri & Goldstein Khatri and Goldstein (2015a,b) used a similar model, based on a simple biophysically motivated genotype-phenotype map of a single transcription factor binding to DNA; they found that the nature and rapidity of the growth of DMIs between two allopatric populations strongly depended on the product of the effective population size and the scale of fitness of the stabilising landscape; for populations that are small compared to the inverse of the scale of fitness, DMIs arise more quickly. This arises as there are exponentially more sequences that bind poorly than well, which means drift pushes common ancestors on average closer to incompatible regions and so less substitutions are needed for incompatibilities to arise in hybrids. In this low population size limit DMIs arise quadratically with divergence time, in agreement with Orr's predictions, however, the underlying mechanism for this power law is very different to Orr's. For populations that are larger than the inverse of this scale of fitness, DMIs take longer to arise as common ancestors have a smaller drift load and require more substitutions for hybrids to reach incompatible regions. However, in addition they arise more slowly, because within a discrete stabilising landscape, as the population size increases the scaled fitness differences becomes larger and more deleterious, giving rise to slower divergence. This is in line with predictions of the nearly neutral theory **??**, where the number of nearly neutral mutations decreases as the population size increases. In this case the growth of DMIs has a characteristic non-power law form, with negative curvature on a log-log plot, indicating that the hybrid binding energies change diffusively. However, real gene regulatory systems are more complex than a single TF binding to DNA, so again the question arises do these predictions hold for more complex gene regulatory systems with more realistic fitness landscape?

Although there has been much progress in understanding evolution in terms of selection, mutation and genetic drift, much of this work has been reliant on phenomenological fitness landscapes, which encompass in a heuristic manner smoothness, epistasis and neutrality Higgs and Derrida (1992); Kauffman and Levin (1987). In recent years, the question of the structure of real fitness landscapes has gained prominence, where the redundancy of the mapping from genotype to phenotype can give rise to non-trivial properties of the evolution of phenotypesFontana (2002); Khatri *et al.* (2009); Hayden *et al.* (2011); Goldstein (2011); Schaper and Louis (2014); Greenbury *et al.* (2014); Manrubia and Cuesta (2015); Greenbury *et al.* (2016). In particular, Khatri et al Khatri *et al.* (2009) introduced a simple genotype-phenotype map for spatial gene expression regulation in development, from which emerged a number of non-trivial features such as a balance between selection and sequence entropy deciding the course of evolution at small population sizes and a partitioning of the effective phenotypic landscape into neutral and selective degrees of freedom; none of these emergent phenomena could be predicted on the basis of purely phenotypic considerations. In this paper, we will use a slightly modified version of the spatial patterning model in Khatri *et al.* (2009), that has explicit sequence representation of each loci, to examine the Dobzhansky-Muller mode of evolution of incompatibilities in allopatry as a function of population size and under stabilising selection in each lineage. Our results show that population size interplays with fitness and sequence entropic effects in a way mirrored by previous results on single TF-DNA binding. In addition, we find that, unlike Orr's prediction, pair-wise interactions between loci dominate the growth of DMIs, suggesting that biophysics provides a stronger constraint on their evolution than the simple combinatorics of pathways between the common ancestor and present day lineages.

Materials and Methods

GP map

The genotype-phenotype map we use was described in detail by Khatri et al.Khatri *et al.* (2009), so here we summarise its basic elements and recapitulate the main results in the Supporting Information, since the methodology used is slightly different. The evolutionary task set for the gene regulation module is to turn an exponentially decaying morphogen gradient across a field of cells in an embryo (M), into a sharp step function profile of a downstream transcription factor TF with its transition at the mid-point of the embryo, as shown in Fig.1. The model for this gene regulation is essentially the model of the genotype-phenotype map, where the binding affinity of the different protein species to themselves or to the regulatory region of TF is determined by matching binary sequences. More specifically, gene regulation is controlled by two non-overlapping binding sites, the promoter P and an adjacent binding site B, together with two protein species, the morphogen M and RNA Polymerase R, where the length of these sequences are ℓ_b. The various binding energies are proportional to the Hamming distance between the relevant sequences, where for protein-DNA binding the cost of a mismatch is ϵ_b and for protein-protein interactions ϵ_g. So for example, $E_{MP} = \epsilon_b \rho(\boldsymbol{b}_M, \boldsymbol{b}_P)$ is the binding energy between the morphogen and the promoter, where ρ is the Hamming distance and $\boldsymbol{b}_M, \boldsymbol{b}_P$ are the morphogen binding sequence and promoter binding sequence respectively.

Given an exponential morphogen concentration profile $[M](x, \alpha)$ as a function of the position of embryonic cells, x, and a fixed concentration $[R]$ of RNAP, in each cell, we follow Shea and Ackers Shea and Ackers (1985) to calculate the TF concentration profile $[TF](x)$, assuming transcription is proportional to the probability of RNAP being bound to the promoter p_{RP}. This probability can be calculated using the canonical ensemble of statistical mechanics, where details can be found in Khatri *et al.* (2009). Since the TF cannot affect its own transcription, the steady state concentration profile is then simply proportional to the probability of RNAP being bound to the promoter: $[TF](x) \propto p_{RP}(\boldsymbol{G}, R, M(x, \alpha))$. where $\boldsymbol{G} = [\boldsymbol{b}_R, \boldsymbol{b}_M, \boldsymbol{b}_P, \boldsymbol{b}_B, \boldsymbol{g}_R, \boldsymbol{g}_M]$ is the genome and is a function of

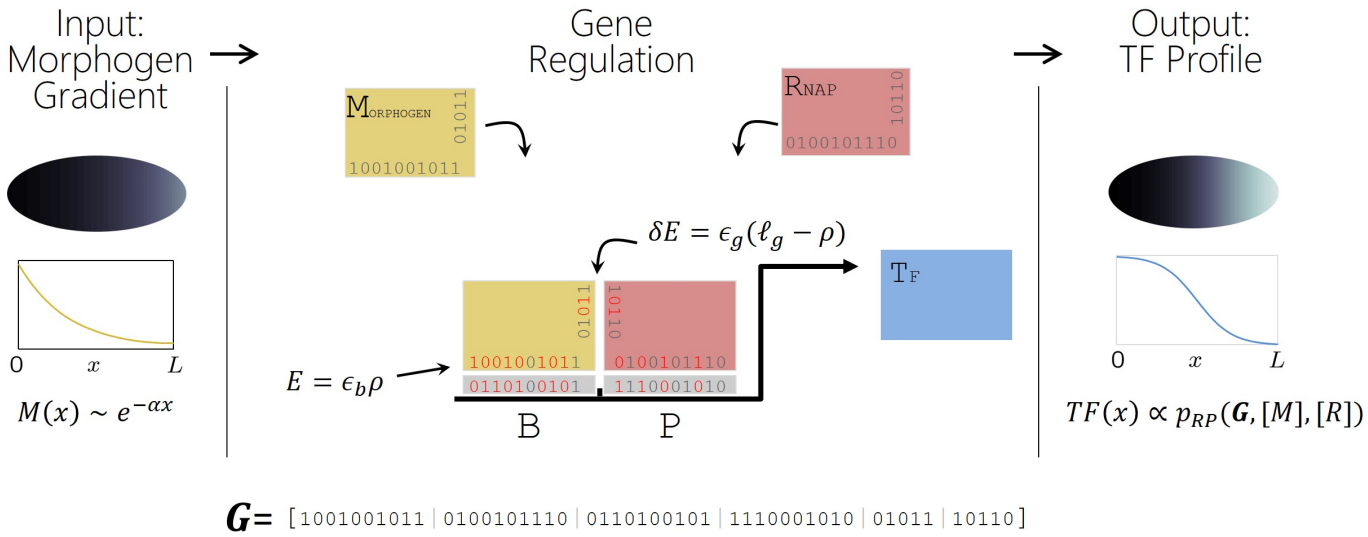

Figure 1 An overview of the genotype-phenotype map. The gene regulatory module has input a morphogen gradient $[M](x)$ across a 1-dimensional embryo of length L and outputs a transcription factor $TF(x)$. Gene regulation of TF using a morphogen and RNAP (R) is controlled in a bottom-up manner, by binding to its regulatory region consisting of a promoter P and adjacent binding site B; E represents binding free energies of proteins to one of the two binding sites of the regulatory region of the transcription factor T, δE are glue free energies to aid in co-operative binding of paired protein complexes. Each energy is calculated by the number of mismatches ρ (Hamming distance), shown in red, between relevant binary sequences, together with mismatch energies ϵ_b and ϵ_g for binding and glue energies respectively. Transcription of T is controlled by the probability of RNAP being bound to the P, p_{RP}.

all the binding and glue sequences from proteins and DNA. The proportionality constant will be given by the ratio of the rate of transcription and translation to the rate of degradation of TF, which is not important in our study, since we are only interested in the shape or contrast of $[TF](x)$ that can be achieved.

Monte Carlo Scheme

We use a kinetic Monte Carlo scheme to simulate a Wright-Fisher evolutionary process for the genome $\boldsymbol{G}$ and α on each lineage, as detailed in Khatri and Goldstein (2015b). In particular, the rate of fixation of one-step mutants are calculated based on Kimura's probability of fixation Kimura (1962), where we are assuming a regime of small effective population size ($n\mu_0 N_e \ll 1$, where $n = 50$ is the number of coding base-pairs in the genome and μ_0 is the base-pair mutation rate and N_e is effective size of the haploid asexual population). Here, we determine the goodness of the spatial gene regulation, from the resulting concentration profile $[TF](x)$ by mapping to a Malthusian fitness by use of a functional that promotes expression of the TF in the anterior half, whilst penalising expression in the posterior half, with truncation selection below critical a value W^*. The functional $\mathcal{W}$ is:

$$\mathcal{W}[[TF](x)] = \frac{\int_0^{L/2}[TF](x)\mathrm{d}x - \int_{L/2}^{L}[TF](x)\mathrm{d}x}{\frac{L}{2}\max_x\{[TF](x)\}}. \quad (1)$$

We then map this to a Malthusian fitness as follows:

$$F = \begin{cases} \kappa_F \ln(\mathcal{W}) & \text{if } \mathcal{W} > W^* \\ -\infty & \text{if } \mathcal{W} < W^* \end{cases} \quad (2)$$

where κ_F is the strength of selection for the trait represented by the spatial patterning process. Note that although here the exact form of the fitness is slightly different to the one used in Khatri *et al.* (2009), the qualitative behaviour is the same as shown in the Supporting Information.

Speciation Simulations

Starting from a random initial genome and fixed population size N_e, the spatial patterning simulation is run for 100,000 substitutions, in order to effectively equilibrate the system (typically 10,000 substitutions are required to adapt to an ensemble of fit states). For each replicate allopatric speciation run, two simulations are performed from the same common ancestor, each using the same fitness function Eqn.2. For each simulation pair, the common ancestor is drawn from the equilibrium distribution taking the end state of the sequences of the long simulation and running it further for a further 100 substitutions. Although, non-neutral substitutions can fix on each line, since each line has the same environment we consider this situation to be neutral with respect to selection on each line.

For simplicity, we assume that our genome is composed of 4 loci: 1) RNAP (R), 2) Morphogen (M), 3) Regulatory region of TF (T) and 4) the morphogen gradient steepness α, and we form hybrids between the two lines assuming independent re-assortment of these loci and complete linkage within each loci. We define a hybrid genotype by a 4 digit string where each digit corresponds to one of the loci defined above and takes one of two values, which correspond to the allele from the 1st line or 2nd line; for example, the hybrid rMTa corresponds to R loci having an allele from the 1st line, M loci from the 2nd loci, T loci from the 2nd loci and α loci the allele from the 1st loci. Note that the underlying sequence of each hybrid changes as different

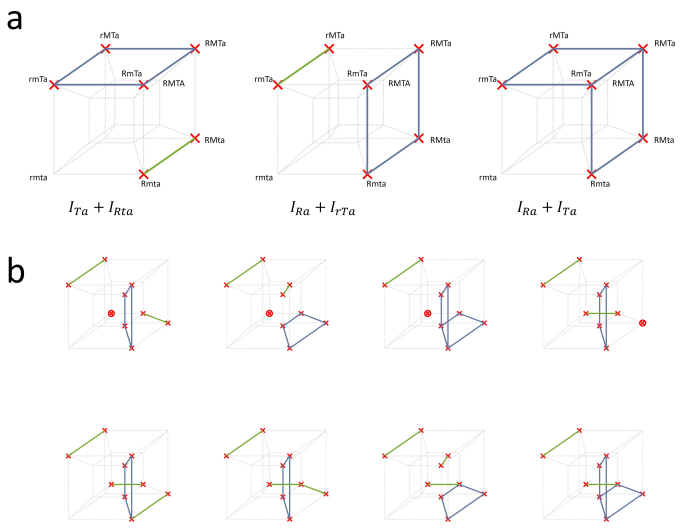

Figure 2 Decomposition of hybrid DMIs on a Boolean hyper 4−cube. Each point on the 4−cube represent each possible hybrid genotype across 4 loci, including the genotype of each parental lineage, where each red cross represents an incompatible hybrid genotype. As shown in a) the pattern of DMIs can be explained by different combinations of fundamental types of DMIs, where a blue square or face identify a subspace of genotypes that correspond to a 2−point DMI and a single green edge of line corresponds to a 3−point DMI. b) A more complicated pattern of hybrid DMIs and their decomposition into fundamental types, where a red dot corresponds to a single isolated 4−point DMI.

substitutions are accepted in each line; the notation only refers to alleles fixed at any point in time. Under these assumption we can then look at how the average fitness and average number of incompatibilities varies as the number of substitutions that separate two lines increases. As substitutions in α are quite common due to a small curvature of its landscape, we only count substitutions in α when it makes a transition between the two basins of attraction mentioned above and discussed in Khatri *et al.* (2009).

The value chosen for W^* is somewhat arbitrary, the larger W^* the more quickly incompatibilities will arise; here we choose $W^* = 0.2$ ($F^* = \kappa_F \ln W^* \approx -0.0016$), which gives a good balance between the number of incompatibilities on a given timescale and realism for their actual inviability; arguably $W^* = 0$, but this requires quite lengthy simulations in order for incompatibilities to arise.

Decomposing DMIs

Examining the number of DMIs that arise across different hybrid genotypes is not the most fundamental representation of the incompatibilities, since for example, if a particular hybrid RmTa, where the 1st and 3rd loci take alleles from line 2 and the 2nd and 4th take alleles from line 1, is found to be incompatible, it is not known whether this incompatibility arises from a pairwise interaction between R and m or between m and T, or a triplet interaction between R,m and T or all of these concurrently. Ideally we would like to be able to decompose fitness contributions into n-point interactions between loci, for example, using a Fourier series on a Boolean hypercube Weinberger (1991); Neher and Shraiman (2011); however, such decompositions suffers from a problem that the terms involving interactions between 3 or more loci are difficult to interpret physically. A more explicit representation of the form $F(\boldsymbol{g}) = F_0 + \sum_i F_i(g_i) + \sum_{ij} F_{ij}(g_i, g_j) + \sum_{ijk} F_{ijk}(g_i, g_j, g_k) + F_{1234}(g_1, g_2, g_3, g_4)$ is not possible as the problem is hugely underdetermined; given that there are only $2^4 - 2 = 14$ possible hybrids (not including the well-adapted genotypes of line 1 and line 2) and a total of $I_{max} = 3^L + 1 - 2^{L+1} = 50$, for $L = 4$ loci [1], different interactions it is not possible to determine unambiguously which interactions are responsible for a particular set of DMIs across all hybrids. The approach we take is look at patterns of DMIs across all hybrid genotypes and construct the most parsimonious fundamental interactions that could give rise to this pattern of DMIs across the hybrid genotypes. Here we define parsimonious to be, all other things being equal, the minimum number of possible interactions needed to explain the observed pattern of DMIs, which from a Bayesian perspective would have the smallest Occam factors MacKay (2007).

If we consider a pair-wise incompatibility say T ↮ a, which we denote I_{Ta}, any hybrid-genotype that contains the alleles Ta must be a DMI; this defines a $2D$ subspace (or face) of 4 DMIs across loci R and M on a $4D$ Boolean cube $\{\mathsf{rmTa}, \mathsf{rMTa}, \mathsf{RmTa}, \mathsf{RMTa}\}$. Similarly, a 3-point DMI I_{mtA} defines an edge of 2 DMIs along the $1D$ subspace of loci 4: $\{\mathsf{rmtA}, \mathsf{RmtA}\}$. Each 4-point DMI takes up a single point in the hybrid-genotype space. The approach is then to find all combinations of possible DMIs that could explain the pattern of incompatibilities across the hybrids, but consider only those with the least number of total DMIs. Of this subset of minimal DMIs, we assume there is no a priori reason each subset should not be given equal weighting and so then our measure of the number of DMIs of each type is the average number of 2-point, 3-point and 4-point DMIs. For example, in Fig.2a, the hybrids may have a pattern of DMIs as shown by the red crosses; the most parsimonious representation are 3 equally weighted possibilities: 1) $I_{\mathsf{Ta}} + I_{\mathsf{Rta}}$, 2) $I_{\mathsf{Ra}} + I_{\mathsf{rTa}}$, 3) $I_{\mathsf{Ta}} + I_{\mathsf{Ra}}$. If this pattern of hybrid incompatibilities were to arise it would be decomposed into $n_{Ta} = 2/3$, $n_{Ra} = 2/3$, $n_{Rta} = 1/3$, $n_{rTa} = 1/3$, so the total number of 2-point DMIs $n_2 = 4/3$ and 3-point DMIs $n_3 = 2/3$, corresponding to a total number of $n = n_2 + n_3 = 2$ DMIs, which is the minimum number of DMIs needed to explain the pattern of hybrid incompatibilities shown in Fig.2a. In Fig.2b is shown a more complicated pattern of hybrid DMIs and the parsimonious decomposition into k-point DMIs.

[1] The total number of n−point DMIs is $(2^n - 2)\binom{L}{n}$, as there are $\binom{L}{n}$ combinations of n loci amongst L total loci and then considering a binary choice of alleles across both lines, there are a total of 2^n allelic combinations or states, 2 of which are the fit allelic combinations where all alleles come from one lineage or the other giving $2^n - 2$. For example, between each pair of loci there are $2^2 - 2 = 2$ mismatching combinations of alleles (e.g. rM and Rm) that could give DMIs and $\binom{L}{2} = L(L-1)/2 = 6$ pairwise interactions. A similar argument would give a total of 24 3-point DMIs as there are $2^3 - 2 = 6$ mismatching combinations of alleles at 3 loci (e.g., excluding rmt and RMT) and $\binom{L}{3} = 4$ 3-point interactions and similarly, $14\binom{L}{4} = 14$ for 4-point interactions. In total, the max number of DMIs is $I_{max} = \sum_{n=2}^{L}(2^n - 2)\binom{L}{n} = 3^L + 1 - 2^{L+1}$, which for $L = 4$ loci is $I_{max} = 50$.

Results

Evolutionary properties of genotype-phenotype map on each lineage

The properties of this genotype-phenotype map have been previously explored Khatri *et al.* (2009). As the some small details of the evolutionary dynamics have changed here, we show in the Supporting Information that the basic pertinent findings from Khatri *et al.* (2009) are recapitulated. An important property of this genotype-phenotype map is that only a single mechanism of patterning is found, which is that RNAP (R) binds with intermediate affinity to the promoter (P), but through a high affinity protein-protein interaction with the morphogen (M), the morphogen binds to the first binding site (B) only above a critical morphogen concentration, thereby giving a spatial switch once the morphogen falls below this concentration; evolution then fine tunes the relationship between the binding energies (E's), glue energies (δE's) and the steepness of the morphogen gradient α to turn off transcription at the mid-point of the embryo. Despite a single global solution there are many different combinations of the binding and glue energies and α that give good patterning, and for each of these many underlying genotypes ($\boldsymbol{G}$). A further key property is that despite this redundancy some energy phenotypes such as $E_{MB}, \delta E_{RM}$ and E_{RP} are under strong stabilising selection, whilst the remaining energy phenotypes, $E_{RB}, E_{MP}, \delta E_{RR}$ and δE_{MM} are almost neutral and under very weak stabilising selection (see Supporting Information and Khatri *et al.* (2009)). Finally, at large population sizes it is found that the evolutionary dynamics exhibits what is known as quenched-disorder in statistical physics, where energy phenotypes are that less constrained take different *random* values between independent evolutionary runs, which indicates there is an underlying roughness to the fitness landscape Khatri *et al.* (2009).

Different hybrids genotypes have different growth rates of DMIs

In Fig.3, we plot a typical time series, for scaled population sizes of $2N\kappa_F = 1$ and $2N\kappa_F = 10$, of how fitness of two different hybrids (Rmta a & b and RMtA c & d) changes over time t separating a pair of lines together with the threshold in fitness (dashed black line) used later to count DMIs. We see that the fitness of hyrbids generally decreases in a stochastic fashion; when the fitness of a hybrid drops below the threshold F^*, a DMI arises and is indicated by the disappearance of the fitness line ($F = -\infty$) for that hybrid and so as can be seen from Fig.3, at any given time only a subset of all possible hybrids might be incompatible. We also see that as the fitness of hybrids is stochastic, DMIs that arise do not stay, as one might expect within the Orr framework Orr (1995); Orr and Turelli (2001). A further observation is that for smaller scaled population sizes the common ancestor fitness is lower and incompatibilities appear to arise more quickly; this is a consistent with previous studies of a more simple genotype-phenotype map of a transcription factor binding a single binding site Khatri and Goldstein (2015a,b), where the smaller populations have a larger genetic drift load and so common ancestors are more likely to be closer to an inviable binding threshold.

To examine the overall trends in the number of incompatibilities for each hybrid type, as a function of divergence time μt, we plot the average number of DMIs in Fig.4, where we have averaged over pairs of complementary genotypes (e.g. RmtA and rMTa), which behave the same on average as each line is exploring the same fitness landscape independently (not shown). As denoted in the figure, we distinguish hybrid-genotypes based on the types of potential mismatch: R-type is characterised by a mismatch of the R loci with M and T loci, M-type is a mismatch of M with R and T loci, T-type is a mismatch of T with R and M and α-type is a mismatch of the α loci with the rest of the loci. In Fig.4, we first note that each hybrid-genotype behaves in a different way in a population size dependent manner. In general, there is an initial growth in the average number of DMIs as the time separating lines increases, followed by a slowing down of the growth or a plateau. It is also clear that as the population size is increased DMIs take longer to arise, which is also consistent with previous work on the dynamics of incompatibilities due to transcription factor DNA binding Khatri and Goldstein (2015b), where this is caused by a slowing of the substitution rate as in a stabilising discrete fitness landscape more deleterious changes are needed for any evolutionary change. We also see that for small population sizes the initial growth of DMIs is power law and approximately quadratic, as predicted by Orr Orr (1995); Orr and Turelli (2001), but as argued previously Khatri and Goldstein (2015b) the underlying mechanism is very different. On the other hand for large population sizes there is no clear power law, which is again consistent with previous simulations Khatri and Goldstein (2015b) and also theoretical calculations Khatri and Goldstein (2015a) that predict that as common ancestor populations are further away, the growth of DMIs follows a diffusive law, which has a characteristic negative curvature on a log-log plot.

However, in addition to these general trends, we see that the growth of DMIs is different for different hybrids; for small population sizes ($2N\kappa_F = 0.1$ & $2N\kappa_F = 1$), M-type, T-type and R-type dominate the growth of DMIs, in this order and with only a small difference between them, whilst α-type arise far more slowly; we might expect this since substitutions in α only tend to shift the pattern away from the mid-point of the embryo, which with the model of fitness defined in Eqn.2 only moderately affects fitness. As the population size increases, and at small times, we see that initially M-type DMIs arise more slowly relative to T-types and R-types, but at longer times there is a cross-over where M-type DMIs dominate R-type; although the simulations do not run out to sufficiently long times for the largest population sizes, this cross-over appears to move to longer times at increasing population sizes.

How can we understand this general behaviour? It is clear from Eqns. 1&2 that good patterning or fitness is only dependent on the binding and glue energy phenotypes (as well as α) and so this in general requires co-evolution of the relevant sequences to maintain these energies within certain constraints; e.g. the binding energy of R to the promoter of T, E_{RP} mustn't be too strong and so on each line the sequences will co-evolve to maintain this constraint. From previous work Khatri and Goldstein (2015a,b) we expect that it is not only the population size that determines how quickly hybrid incompatibilities arise, but its product with the strength of selection κ_β maintaining a particular interaction; so the energy E_{MB} is under very strong selection for strong binding of the morphogen to the 1st binding site and we would expect incompatibilities due to this interaction to arise more slowly when $2N\kappa_{MB} \gg 1$.

Decomposition of DMIs

As discussed in the model section the DMIs shown in Fig.4 will have contributions from many different fundamental incompatibility types, which can be 2-point, 3-point and 4-point

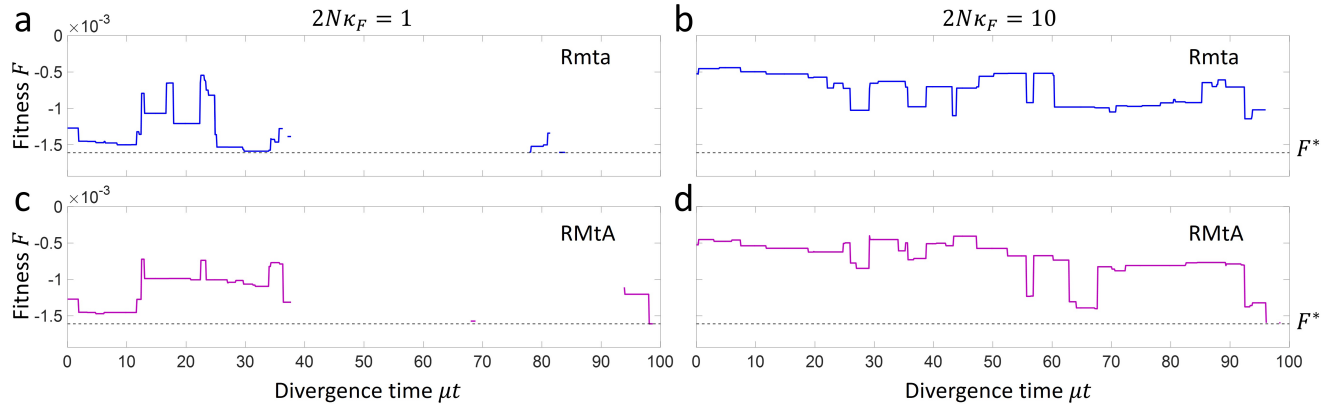

Figure 3 Plot of the times series of two hybrids Rmta (a & b) and $RMtA$ (c & d) at population sizes of $2N\kappa_F = 1$ (a & c) and $2N\kappa_F = 10$ (b & d).

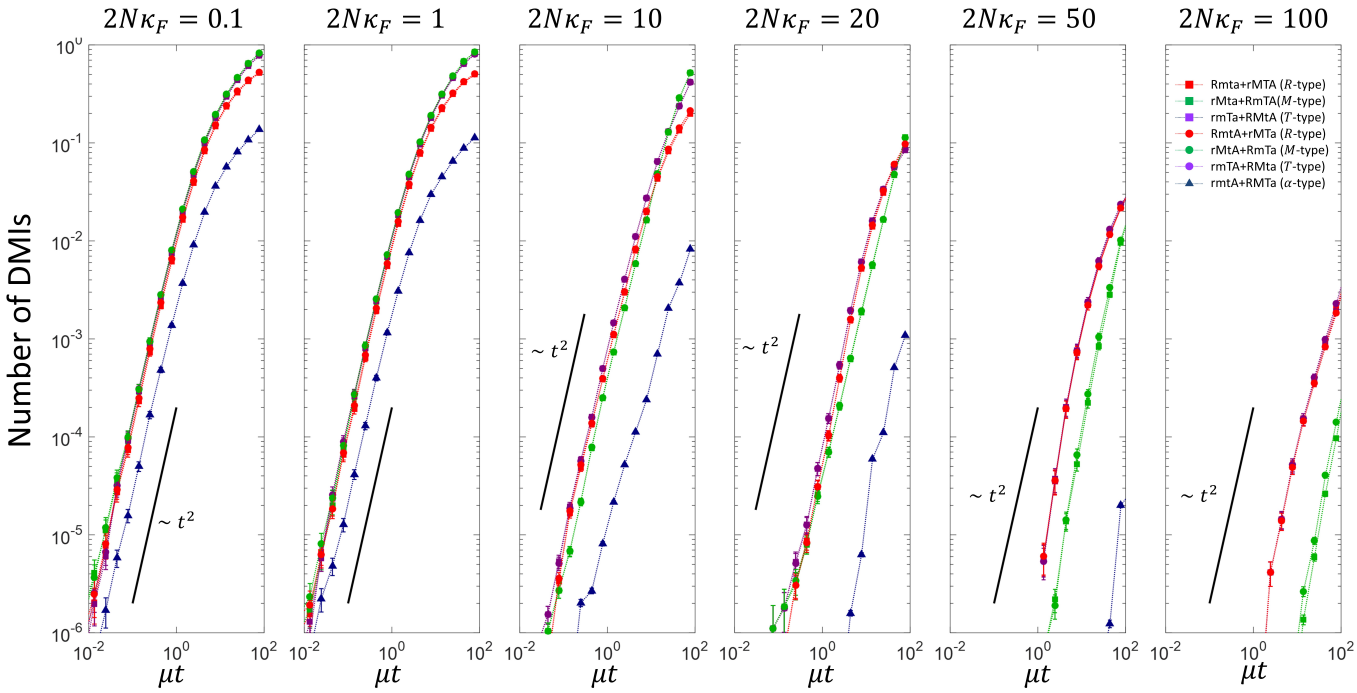

Figure 4 Plot of the number of DMIs for each hybrid genotype since divergence for different scaled population sizes. Complementary hybrid genotypes are summed over, since their behaviour should on average be the same, as there is the same stabilising selective pressure on each lineage.

in nature. Using the method described above to decompose DMIs into fundamental types, we plot the total number of each type of DMI versus divergence time in Fig.5, where the panels correspond to different scaled effective population sizes from $2N\kappa_F = 0.1$ to $2N\kappa_F = 100$. We see clearly that pair-wise DMIs are domininant at all population sizes and divergence times, though the difference is diminished at larger population sizes. In addition, for small population sizes, the relative proportion of 2-point to 3- and 4-point DMIs passes through a minimum at intermediate divergence times. These results show that contrary to the prediction of Orr that higher order DMIs should be easier to evolve, higher order DMIs evolve more slowly and are in smaller number compared to pair-wise DMIs.

As mentioned in the introduction the Orr model also predicts that n-point DMIs should increase as $\sim t^n$. Here, we find that for small population sizes 2-point, 3-point and 4-point DMIs all increase as a power law at small times, indicated by a straight line on a log-log plot, with a larger exponent for 3-point and 4-point DMIs. To more quantitatively assess the exponent, we fit the data for $2N\kappa_F \leq 20$ using the phenomenological equation:

$$I(t) = \frac{I_0 t}{T+t}\left(1 - \exp\left(-t/\tau\right)^{\gamma-1}\right) \tag{3}$$

which has the asymptotic form of $I(t) \sim t^\gamma$ for $t \ll \tau$ and $t \ll T$ and an opposite limit of $I(t \to \infty) = I_0$. We see that for the total number of DMIs and for 2-point, 3-point and 4-point DMIs, this form fits the data well at intermediate and small population sizes. We tabulate the power law exponent derived from these fits in Table 1. We see that the total number of DMIs and 2-point DMIs have a power law exponent close to $\gamma = 2$, which is consistent with the Orr model and with Fig.4 which shows a similar power law, further showing that 2-point DMIs are dominant in determining the growth of hybrid incompatibilities. However, the higher order incompatibilities do not quite follow the Orr prediction, where n-point DMIs have an exponent $\gamma = n$, although 4-point DMIs have a larger exponent than 3-point DMIs; 3-point DMIs have an exponent that varies between $\gamma = 2$ and $\gamma = 3$, while 4-point DMIs have an exponent between $\gamma = 3$ to $\gamma = 3.5$. In all these cases an alternative model for the power law behaviour, as argued in Khatri and Goldstein (2015b), is that at small population sizes, where genetic drift is dominant and there is a large drift load, common ancestor populations are poised at the incompatibility boundary (truncation selection threshold) and the growth of DMIs at short times is due to how likely a few critical substitutions are to arrive very quickly, which is given by a Poisson process; if the critical number of substitutions is K^* then for short times we would expect $P_I(t) \sim (\mu t)^{K^*}$ and so given that at least n substitutions are needed for a n-point incompatibility, we would expect $K^* \geq n$. It is possible the inconsistency here could be resolved by more accurate measurement of the power law, by exploring simulations at even shorter times, using a larger number of replicate simulations (here there are 10^6 replicate simulations at short times).

At larger population sizes ($2N\kappa_F \geq 50$), Fig.5 shows that there is no clear power law and instead there is a negative curvature in the growth of DMIs on a log-log plot. This is consistent with a model of DMI growth where high fitness corresponds to high binding affinity, so that the common ancestor distribution is peaked away from the inviability boundary, which would arise with large populations that have a small drift load, meaning that hybrids diffuse to the boundary to give rise to DMIs; one such analytically tractable model was investigated in Khatri and Goldstein (2015a) and predicted that the number of DMIs is a complementary error function, which has an asymptotic form $I(t) \sim \frac{\sqrt{4\mu t}}{K^*} e^{-(K^*)^2/4\mu t}$, which due to the essential singularity as $t \to 0$ has the property of negative curvature on a log-log plot. However, this form does not fit the simulation data well (not shown). Given the multidimensional nature of this spatial patterning model, it is possible that we need to consider the analogous result to the effective one-dimensional diffusion studied in Khatri and Goldstein (2015a), which results in a multidimensional generalisation of the error function Brown (1963), where in n dimensions $\mathrm{erf}_n(x) = \Gamma(x^2, n/2)/\Gamma(n/2)$, so that the number of DMIs has asymptotic form $I(t) \sim (\frac{4\mu t}{K^*})^{1-\frac{n}{2}} e^{-(K^*)^2/4\mu t}$; this again, however, does not fit the data in Fig.5 well. A functional form that is a good fit for the data at large populations sizes is

$$I(t) = \frac{\sqrt{4(\mu t)^\beta}}{K^*} e^{-(K^*)^2/4(\mu t)^\beta}, \tag{4}$$

which arises when considering fractional Brownian processes with exponent β; normal diffusion or Brownian motion arises when $\beta = 1$, while $\beta < 1$ corresponds to *sub*diffusive behaviour, while $\beta > 1$ is *super*diffusive. It is clear by examining the exponent β in Table 2 from fits of the data in Fig.5 at large population size ($2N\kappa_F \geq 50$) that the DMIs arise as a result of a subdiffusive process, where 2−point, 3-point and 4-point DMIs have an exponent $\beta \approx 1/3$ for $2N\kappa_F = 50$ and $\beta \approx 1/4$ for $2N\kappa_F = 100$. The most likely mechanism that would give rise to subdiffusive behaviour is a broad spectrum of times between substitutions; even though in the simulations the kinetic Monte carlo scheme is based on a Poisson process for a given genotypic state $\boldsymbol{G}$, the distribution of rates could vary significantly as populations explore the fitness landscape. This would be consistent with the results in Khatri *et al.* (2009), which reveal the underlying fitness landscape of this spatial patterning genotype-phenotype map to be rough, which could lead to broad distribution of substitution rates in each lineage and effective subdiffusive behaviour of the hybrids Bertin and Bouchaud (2003). Finally as expected the average number of substitutions needed at large population sizes is large, irrespective of the value of n, with values of K^* ranging from 6 to 9.

2-point DMIs In Fig.6 we have plotted the number of 2-point DMIs of each type, where for example, I_{mt} is a 2-point DMI caused by an incompatibility between the M locus and T locus. For each type of 2-point incompatibility there are 2 binding energy traits that could contribute. So increases in I_{mt} could be due to an incompatibility in a hybrid in E_{MB} or E_{MP}; in this case, as the binding energy E_{MP} is almost neutral Khatri *et al.* (2009), we would would expect incompatibilities to arise predominantly

$2N\kappa_F$	0.1	1	10	20
Total	1.94 ± 0.05	1.99 ± 0.05	1.95 ± 0.05	2.00 ± 0.12
2-point	1.93 ± 0.05	1.98 ± 0.06	1.84 ± 0.06	1.97 ± 0.13
3-point	2.66 ± 0.13	2.81 ± 0.17	2.76 ± 0.24	2.14 ± 0.15
4-point	3.17 ± 0.19	3.43 ± 0.41	3.14 ± 0.24	2.93 ± 0.19

Table 1 Table of values of the exponent γ characterising the power law of growth of DMIs at short times and small scaled population sizes.

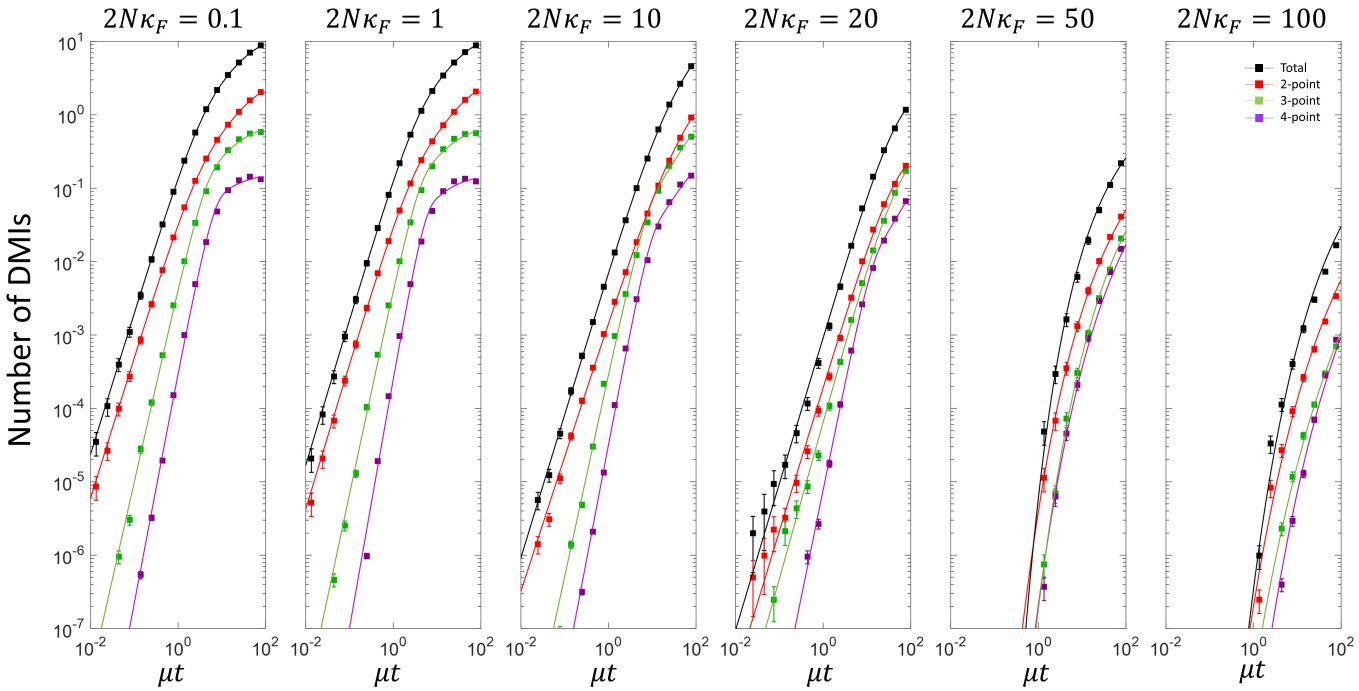

Figure 5 Plot of the total number of DMIs vs divergence time, together with their decomposition into the total number of $2-$point, $3-$point, $4-$point DMIs, for various scaled populations sizes. For $2N\kappa_F \leq 20$ the solid lines correspond to fits of the simulation data to Eqn.3, while for $2N\kappa_F \geq 50$ correspond to fits to Eqn.4.

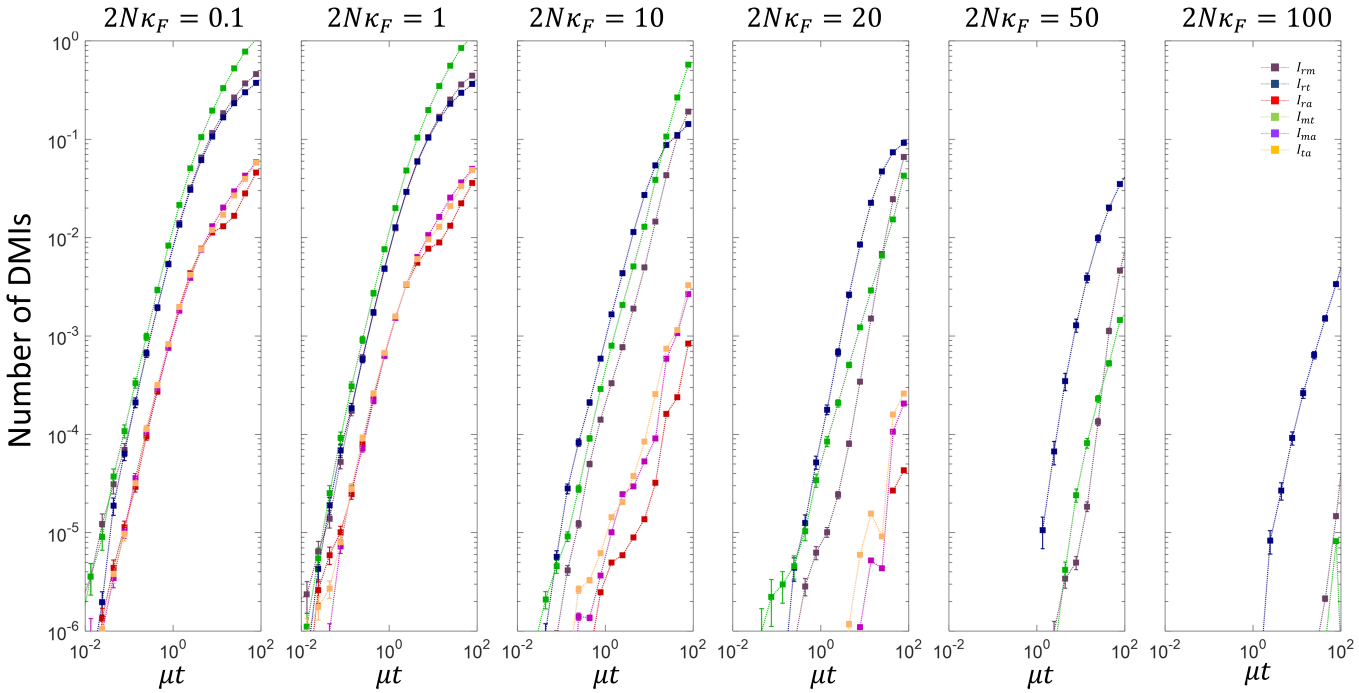

Figure 6 Plot of the spectrum $2-$point DMIs vs divergence time for different scaled population sizes.

$2N\kappa_F$		50	100
Total	β	0.47 ± 0.03	0.33 ± 0.02
	K^*	6.58 ± 0.12	7.37 ± 0.26
2-point	β	0.32 ± 0.01	0.25 ± 0.01
	K^*	6.71 ± 0.07	7.51 ± 0.18
3-point	β	0.33 ± 0.01	0.22 ± 0.01
	K^*	7.53 ± 0.10	8.09 ± 0.27
4-point	β	0.31 ± 0.02	0.25 ± 0.02
	K^*	7.50 ± 0.28	8.66 ± 0.39

Table 2 Table of values of the parameters characterising the sub-diffusive growth of DMIs for large scaled population sizes; $\beta = 1$ corresponds to normal diffusive motion, $\beta < 1$ to sub-diffusion and $\beta > 1$ super-diffusion, while K^* corresponds roughly to the number of substitutions required to reach the invaible region.

from E_{MB}. Similarly, we expect I_{rt} to be dominated by E_{RP} and not E_{RB} and I_{rm} dominated by δE_{RM} and not δE_{MM} or δE_{RR}. When it comes to incompatibilities involving the α-locus, there is no clear phenotypic trait that can be identified and as we will see the analysis of these DMIs will not be so clear.

Examining Fig.6, we see at small population sizes $2N\kappa_F \leq 1$, that all the DMIs grow approximately quadratically at short times with a saturating form at long times, as seen in Fig.5. In addition, we see that when population sizes are small, the number of DMIs for a particular pair wise interaction correlates with the strength of selection on that trait; for example, the binding energy trait E_{MB}, which has the strongest selective constraint, gives rise to the most number of DMIs (I_{mt}) at all times; the next most critical energy trait in terms of selective constraint is δE_{RM}, which has the next highest number of DMIs (I_{rm}). Although, this observation would appear to be intuitive, interpreting it in light the results of a simple model of transcription factor DNA binding Khatri and Goldstein (2015b) are not straightforward; the results of this work suggest that for small population sizes, the rate that incompatibilities arise decreases with increasing strength of selection and decreasing sequence length, so it is possible these two effects could confound each other. Here for example, E_{MB} has a stronger selective constraint compared to δE_{RM}, but a longer sequence length, as each DNA-protein interaction interface has 10 binary digits versus each protein-protein interaction interface that has 5. This sequence length effect arises, simply as a result of the fact that the overall substitution rate of a binding energy is proportional to the number of nucleotides that code for it, and so whilst each substitution will have same phenotypic effect, the hybrid binding energy changes more quickly for longer sequences, as substitutions are more frequent. In addition, it is not obvious how to directly map this single binding interface model onto the multidimensional situation here; one possibility is that the effect of the fitness threshold for truncation selection impacts on each pair-wise interaction differently, for example, where the *average* number of substitutions, that affect the binding energy trait E_{MB}, needed to give rise to an incompatibility I_{mt} is smaller than for other binding energy traits.

However, as the scaled population size increases, we see that the time for I_{mt} incompatibilities to arise sharply increases, while the time for I_{rm} increases less rapidly and I_{rt} even less rapidly. This is consistent with the simple model of transcription factor DNA binding Khatri and Goldstein (2015b) and as observed with the hybrid DMIs in Fig.4, as E_{MB}, which contributes most to I_{mt} is under the greatest selection pressure and so as the population size changes these should change most rapidly. We see that for large population sizes, it is not the phenotypic traits under the strongest selection that give rise to significant DMIs at short times, but those under a weaker selective constraint; traits under weaker selection will be affected more by the sequence entropic pressure for poorer binding affinities and so the common ancestor is more likely to be closer to the inviability boundary. However, if a trait is effectively neutral, i.e. that selection is sufficiently weak that for no values that the trait can take do incompatibilities arise then these will not give rise to incompatibilities; the energy traits E_{RB}, E_{MP}, δE_{RR} and δE_{MM} have this property, as is evident by examining their marginal distribution functions which follow the neutral expectation Khatri *et al.* (2009). The 2-point incompatibilities involving the α locus are more difficult to interpret, since there is no clear trait in the patterning model associated with an interaction solely between the α locus and R, M, or T loci; if α was resolved into a sequence for a protease and it's interaction with a 3rd sequence of the M loci, in addition to the binding and glue sequences, then the value of α itself would be a trait determined by a pair-wise interaction between this 3rd sequence of M and the protease loci, but the current model does not include this feature. The most identifiable phenotype associated with α is the position of the mid-point of the embryo, but this trait involves a co-evolution of E_{MB} and α and so represents a 3-point interaction between M, T and α loci, which will be discussed below. It is likely that the 2-point DMIs involving α are therefore false and a consequence of the parsimonious DMI decomposition method used, which assumes an equal prior on all possible DMIs that have a minimum number of DMI types. This could be rectified by having a zero-prior on all pair-wise incompatibilities involving the α locus; however, here we have not implemented this as 2-point DMIs involving α are typically an order of magnitude smaller than the other DMIs.

3-point & 4-point DMIs In Fig.7, we have plotted the $3-$point DMIs as a function of divergence time μt, where the panels from left to right represent increasing scaled population size. We see that for small population sizes, $3-$point DMIs between the R, M and T loci dominate at all times and in particular that the different types of DMIs of this type are all roughly equal, $I_{Rmt} \approx I_{rMt} \approx I_{rmT}$. In addition, we see that all other DMIs arise more slowly and each of the 9 other types of $3-$point DMIs are all again approximately equal. However, at larger population sizes this degeneracy is lifted amongst the different types of DMIs and different $3-$point DMIs grow at different rates. How can we understand this general behaviour?

The patterning solution found in these evolutionary simulations involves the morphogen binding strongly to the first binding site recruiting RNAP to bind to the promotor to turn on transcription, through a high affinity interaction between the morphogen and RNAP; the spatial position along the length of the embryo where the transcription switches from on to off is controlled by an interaction with the steepness of the morphogen gradient α. Given this, incompatibilities between R, M and T loci could arise through a $3-$point interaction where the R loci interacts with the parts of the T and M loci coding for E_{RP} and

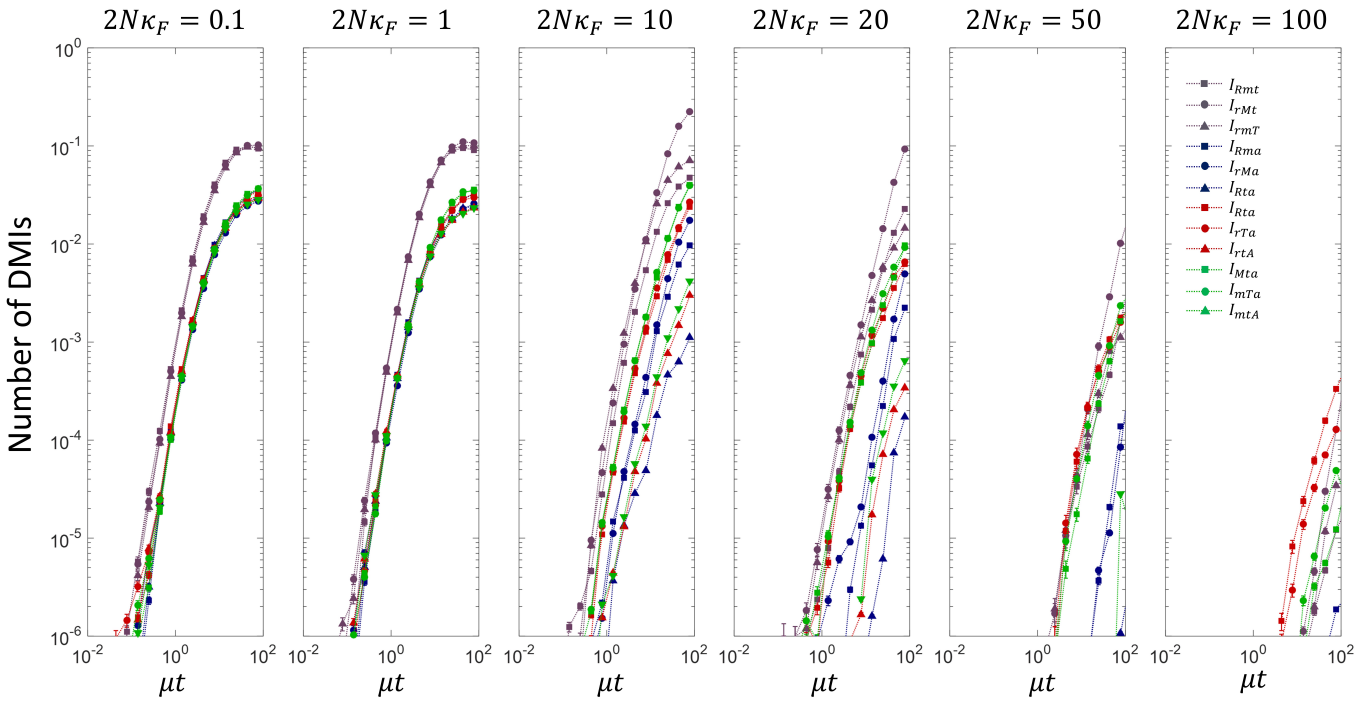

Figure 7 Plot of the spectrum 3−point DMIs vs divergence time for different scaled population sizes.

δE_{RM}, or where M loci interacts with the parts of the T and R loci coding for E_{MB} and δE_{RM}. So in analogy to 2−point DMIs, where a pair of loci give rise a single phenotypic binding energy trait, whose value contributes to fitness, here the triplet of loci, R, M and T, give rise to two binding energy traits, which together contribute to fitness. These two traits will co-evolve to maintain good fitness, balanced by the constraints of sequence entropy on the underlying 3 loci which will diminish at large population sizes. On the other hand 3−point incompatibilities between say M, T and α could arise due to an interaction of the E_{MB} binding energy trait with α; in this model this is subject to a sequence entropy constraint between only two loci. This is true for all the 3−point interactions that involve the α loci. Qualitatively, this then explains the behaviour at low population sizes, as sequence entropy dominates fitness, meaning that the behaviour of the different 3-point DMIs will be dominated by their underlying sequence entropy constraints.

The sequence entropy constraints for the 3−point interactions involving the α loci is straightforward and given by a binomial degeneracy function $\Omega(E) = \binom{\ell}{E/\epsilon}$, so that the sequence entropy function $S(E) = \ln(\Omega)$ is approximately quadratic in E, where here E represents one of the binding energies that interacts with α. However, for the other 3−point interactions that don't involve α, but involve the R, M and T loci, the sequence entropy constraint will be related to a degeneracy function $\Omega(E, \delta E) = \Omega(E)\Omega(\delta E)$, where the joint number of sequences that give E and δE is a product, since these energy traits are coded by different sequences, even though they come from the same loci (the joint number of sequences $\Omega(E_{MB}, E_{MP}) \neq \Omega(E_{MB})\Omega(E_{MP})$ since the protein binding sequence of the morphogen that determines E_{MB} and E_{MP} is the same in this case). Given that the joint number of sequences that give E and δE is a product of two binomial coefficients, the sequence entropy function will approximately be a sum of two quadratic terms $S(E, \delta E) = -\frac{2}{\ell_b}(E/\epsilon_b - \ell_b/2)^2 - \frac{2}{\ell_g}(\delta E/\epsilon_g - \ell_g/2)^2$. At small population sizes, where genetic drift dominates selection, we expect the distribution of common ancestors to be distributed such that they are poised at the incompatibility boundary for E and δE; incompatibilities then arise when substitutions arise that take hybrids across the boundary.

Given that a 3−point DMI between the R, M and T genetic loci corresponds to co-evolution of a pair of binding energy traits, instead of a single binding energy trait for 2−point and 3−point DMIs that involve α, means the fraction of substitutions that lead to incompatibilities versus those that keep the hybrids compatible/fit becomes larger when going from one to two dimensions. This then explains why 3−point DMIs between the R, M and T loci gives rise to incompatibilities more quickly than those involving the α loci, as seen in Fig.7. Also since at small population sizes the only influence that fitness will have is in defining the region of incompatibility for the traits of interest, we see that for each type of DMI there are very little differences in the rate of growth of DMIs.

4−point DMIs correspond to an interaction where all four loci require a particular combination of alleles for good patterning. As previously noted they are of much smaller number compared to 2− and 3−point DMIs, so here, we do not examine these DMIs in detail. However, we note that the 4−point DMIs shown in Fig.8, show a similar pattern as found with 3−point DMIs, where for small scaled population sizes the DMIs tend to cluster, which suggests, as found for 2− and 3−point DMIs, this is due to sequence entropy constraints dominating the growth of DMIs; on the other hand, at large scaled population sizes this degeneracy is lifted and each hybrid has a different growth rate of DMIs, depending on their particular contribution to fitness and how that balances against the sequence entropy constraint.

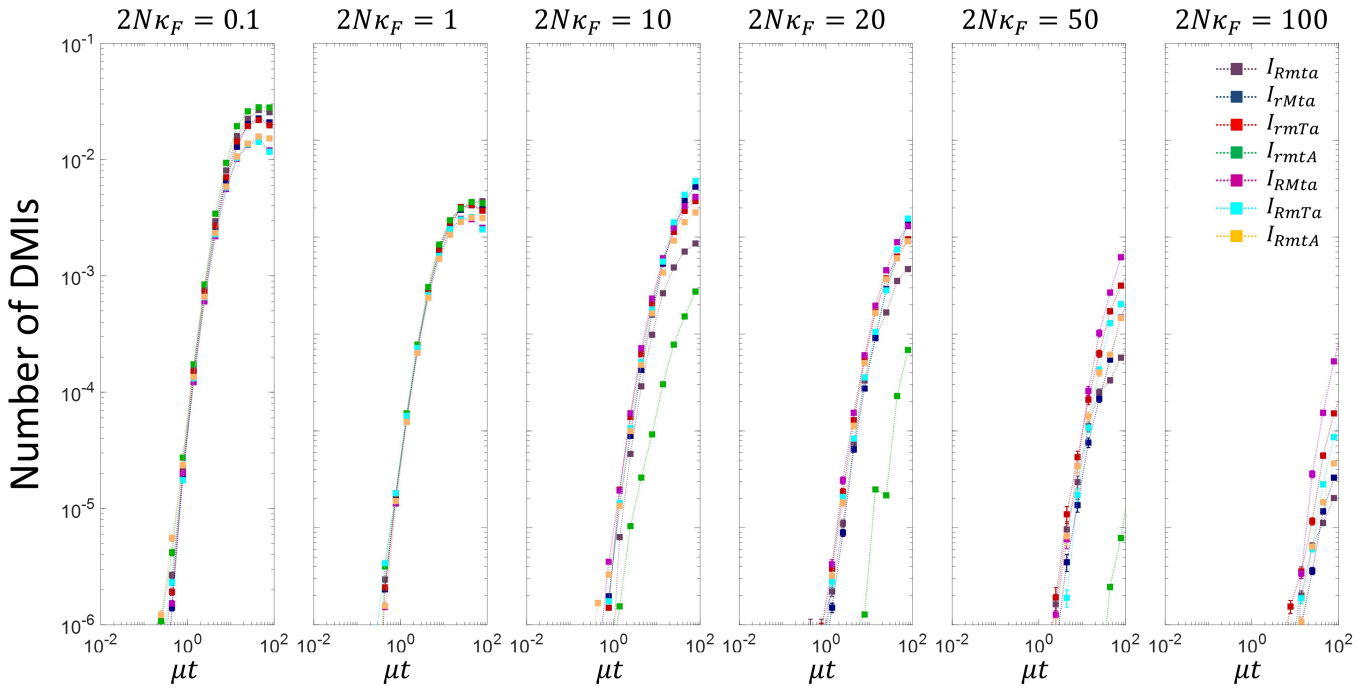

Figure 8 Plot of the spectrum 4−point DMIs vs divergence time for different scaled population sizes.

Discussion

There is still very little understood about the underlying genetic basis that gives rise to reproductive isolation between lineages. Gene expression divergence is thought to be a strong determinant of the differences between species King and Wilson (1975); Wolf *et al.* (2010); Wray (2007); Abzhanov *et al.* (2006); Wittkopp *et al.* (2008) with growing body of evidence for their direct role in speciation, particularly through transcription factors Ting *et al.* (1998); Brideau *et al.* (2006); Mack and Nachman (2016). Here building on previous works which modelled the mechanistic basis and growth of DMIs in models of transcription factor DNA binding Tulchinsky *et al.* (2014b,a); Khatri and Goldstein (2015a,b), we have investigated the growth of DMIs for a simple genotype-phenotype map of gene regulation for spatial patterning in embryonic development, previously studied in Khatri *et al.* (2009). Our results in this more complicated gene regulatory system confirm the basic conclusions from simple models of transcription factor binding Khatri and Goldstein (2015a,b) that 1) as the population size decreases below the inverse of the characteristic scale of fitness ($2N\kappa_F \ll 1$) incompatibilities arise more quickly, 2) they grow in this regime as a quadratic power law with divergence time ($P_I \sim (\mu t)^2$) and 3) for large scaled population sizes incompatibilities arise more slowly with a characteristic negative curvature on a log-log plot indicative of a diffusive process. We note that although we find a quadratic growth of DMIs with divergence time (only at small scaled population sizes), which is as predicted by Orr's framework Orr (1995), the underlying reason is likely to be very different in these models and arises as the common ancestor is likely to be close to the inviable region that gives non-functional binding Khatri and Goldstein (2015b).

In the case of simple models of transcription factor DNA binding Khatri and Goldstein (2015a,b), smaller diverging populations, or traits under weaker selection, were found to develop incompatibilities more quickly, as their common ancestor is already less well adapted due to sequence entropic pressures dominating fitness at smaller population sizes. Here we see that this basic principle that incompatibilities arise more quickly due to a higher drift load of the common ancestor remains valid for a more complicated gene regulatory system. Although this question requires greater empirical attention, there is direct and indirect evidence that smaller populations develop incompatibilities more quickly; for example, the greater species diversity in smaller habitats, such as Hawaii Mayr (1970), the island of Cuba Glor *et al.* (2004) and East African Great Lakes Santos and Salzburger (2012); Owen *et al.* (1990), contrasted with the much slower speciation rate for animals with large ranges or population sizes Mayr (1970, 1954); Rubinoff and Rubinoff (1971); Cooper and Penny (1997). In addition, there is more direct evidence from the net rates of diversification Coyne and Orr (2004) inferred from phylogenetic trees Nee (2001); Barraclough and Nee (2001), which support this population size trend.

The results of this model have also revealed a number of other emergent properties for the growth of hybrid incompatibilities, not obtainable by simply modelling transcription factor DNA binding. For example, for small populations we find clustering of the different behaviours of growth of 3−point DMIs, which can be explained by the different sequence entropy constraints on different binding energies. Also we found that although the growth of DMIs at large population sizes has a characteristic negative curvature on a log-log plot, predicted theoretically by Khatri et al Khatri and Goldstein (2015a), indicating that hybrid traits randomly diffuse, a simple model of diffusion does not fit the simulation data well; instead a model of sub-diffusion that would arise if there are a number of kinetic traps giving a broad distribution of substitution times does fit the data well. This is consistent with the finding that the genotype-phenotype map has a rough fitness landscape, which is only revealed at sufficiently large population sizes Khatri *et al.* (2009).

However, most importantly we find that pair-wise or 2−point

DMIs dominate compared higher order DMIs (3− and 4− point in this model with 4 loci). This is in contrast to Orr's theoretical argument that the fraction of viable paths from the common ancestor to the current day species increases as we consider higher order DMIs Orr (1995). This argument partly rests on the assumption that the number of inviable genotype remains fixed as a larger number of loci are considered, which would seem a very strong assumption. In the same paper Orr also argues that since there are $\binom{L}{n}$ possible $n-$point DMIs (the number of combinations of n loci amongst L loci), so as long as $n < L/2$, we would expect an increase in the number of DMIs as n increases; for $L = 4$ as in this paper, this would suggest $2-$point DMIs are most numerous, but Orr's calculation in fact undercounts the number of DMIs, which as we show above increases as $(2^n - 2)\binom{L}{n}$, in which case $3-$point DMIs, would a priori be more numerous. Our results would then suggest, at least in this simple, but still relatively complex model, that biophysical constraints provide a stronger constraint on the relative number of DMIs of different orders than a purely combinatorial argument would suggest. Although recent results of Weinreich et al Weinreich *et al.* (2013), would seem to contradict our conclusions, their finding of extensive complex epistasis relates to higher order interactions between sites within a single loci, coding for protein stability or enzymatic activity, whereas our work relates to epistasis between multiple loci.

Of course there is an inherent simplicity with our gene regulatory module for spatial patterning, which requires only two proteins to bind to a regulatory region to turn on transcription; a key direction to investigate would be the effect of multiple transcription factors binding to enhancer regions to control gene expression Bintu *et al.* (2005); Spitz and Furlong (2012); Levo and Segal (2014), where there could be a large scope for complex epistasis across many loci coding for a large number of transcription factors. However, as our results show, despite the possibility and a prior expectation of a larger number of triplet interactions, pair-wise interactions dominate; for complex transcriptional control, if pair-wise interactions between proteins, and proteins and DNA dominate, for example in determining the binding affinity of transcriptional complexes, then our conclusions would hold.

Overall, our results point to a basic principle, where developmental system drift or cryptic variation True and Haag (2001); Haag (2014); Gavin-Smyth and Matute (2013), play a key role in speciation; basic body plans or phenotypes are conserved, but co-evolution of the components and loci of complicated gene regulatory networks can change differently in different lineages, giving incompatibilities that grow in allopatry in a manner that is controlled by the drift load of the common ancestor, which is turn determined by a balance between selection pushing populations towards phenotypes of higher fitness and genetic drift pushing them towards phenotypes that are more numerous (higher sequence entropy). In particular, although in principle more complicated regulation should give rise to more complex patterns of epistasis Orr (1995), our findings suggest that more simple, pair-wise, incompatibilities dominate the development of reproductive isolation between allopatric lineages under stabilising selection.

Supporting Information

Acknowledgments

We would like to thank Davis Pollock for initial discussions. RAG was supported by the Medical Research Council under grant U117573805 and BSK by The Francis Crick Institute which receives its core funding from Cancer Research UK, the UK Medical Research Council and the Wellcome Trust.

Literature Cited

Abzhanov, A., W. P. Kuo, C. Hartmann, B. R. Grant, P. R. Grant, and C. J. Tabin, 2006 The calmodulin pathway and evolution of elongated beak morphology in darwin's finches. Nature **442**: 563–567.

Barraclough, T. G. and S. Nee, 2001 Phylogenetics and speciation. Trends in Ecology & Evolution **16**: 391–399.

Bateson, W., 1909 pp. 85–101 in *Darwin and Modern Science*, edited by Seward, A., Cambridge University Press.

Bertin, E. and J.-P. Bouchaud, 2003 Subdiffusion and localization in the one-dimensional trap model. Physical Review E **67**: 026128.

Bintu, L., N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, 2005 Transcriptional regulation by the numbers: models. Curr Opin Genet Dev **15**: 116–124.

Brideau, N. J., H. A. Flores, J. Wang, S. Maheshwari, X. Wang, and D. A. Barbash, 2006 Two dobzhansky-muller genes interact to cause hybrid lethality in drosophila. science **314**: 1292–1295.

Brown, M., 1963 A generalized error function in n dimensions. Technical Report Technical Memorandum No. NMC-TM-63-8, U.S. Navy Missile Center.

Cooper, A. and D. Penny, 1997 Mass survival of birds across the cretaceous-tertiary boundary: molecular evidence. Science **275**: 1109–1113.

Coyne, J. A. and H. A. Orr, 2004 *Speciation*. Sinauer Associates, Inc.

Darwin, C. R., 1859 *The Origin of Species*. J. Murray, London.

Dobzhansky, T., 1936 Studies on hybrid sterility. ii. localization of sterility factors in drosophila pseudoobscura hybrids. Genetics **21**: 113–135.

Fontana, W., 2002 Modelling 'evo-devo' with rna. Bioessays **24**: 1164–1177.

Gavin-Smyth, J. and D. R. Matute, 2013 Embryonic lethality leads to hybrid male inviability in hybrids between drosophila melanogaster and d. santomea. Ecology and evolution **3**: 1580–1589.

Glor, R. E., M. E. Gifford, A. Larson, J. B. Losos, L. R. Schettino, A. R. C. Lara, and T. R. Jackman, 2004 Partial island submergence and speciation in an adaptive radiation: a multilocus analysis of the cuban green anoles. Proceedings of the Royal Society of London B: Biological Sciences **271**: 2257–2265.

Goldstein, R. A., 2011 The evolution and evolutionary consequences of marginal thermostability in proteins. Proteins: Structure, Function, and Bioinformatics **79**: 1396–1407.

Greenbury, S. F., I. G. Johnston, A. A. Louis, and S. E. Ahnert, 2014 A tractable genotype–phenotype map modelling the self-assembly of protein quaternary structure. Journal of The Royal Society Interface **11**: 20140249.

Greenbury, S. F., S. Schaper, S. E. Ahnert, and A. A. Louis, 2016 Genetic correlations greatly increase mutational robustness and can both reduce and enhance evolvability. PLoS Comput Biol **12**: e1004773.

Haag, E. S., 2014 The same but different: worms reveal the pervasiveness of developmental system drift. PLoS Genet **10**: e1004150.

Hayden, E. J., E. Ferrada, and A. Wagner, 2011 Cryptic genetic variation promotes rapid evolutionary adaptation in an rna

enzyme. Nature **474**: 92–95.

Higgs, P. G. and B. Derrida, 1992 Genetic distance and species formation in evolving populations. J Mol Evol **35**: 454–465.

Kauffman, S. and S. Levin, 1987 Towards a general theory of adaptive walks on rugged landscapes. J Theor Biol **128**: 11–45.

Khatri, B. S. and R. A. Goldstein, 2015a A coarse-grained biophysical model of sequence evolution and the population size dependence of the speciation rate. Journal of theoretical biology **378**: 56–64.

Khatri, B. S. and R. A. Goldstein, 2015b Simple biophysical model predicts faster accumulation of hybrid incompatibilities in small populations under stabilizing selection. Genetics **201**: 1525–1537.

Khatri, B. S., T. C. B. McLeish, and R. P. Sear, 2009 Statistical mechanics of convergent evolution in spatial patterning. Proc Natl Acad Sci U S A **106**: 9564–9569.

Kimura, M., 1962 On the probability of fixation of mutant genes in a population. Genetics **47**: 713–719.

King, M. C. and A. C. Wilson, 1975 Evolution at two levels in humans and chimpanzees. Science **188**: 107–116.

Levo, M. and E. Segal, 2014 In pursuit of design principles of regulatory sequences. Nature Reviews Genetics **15**: 453–468.

Mack, K. L. and M. W. Nachman, 2016 Gene regulation and speciation. Trends in Genetics .

MacKay, D. J. C., 2007 IV, p. 343 in *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press.

Manrubia, S. and J. A. Cuesta, 2015 Evolution on neutral networks accelerates the ticking rate of the molecular clock. Journal of The Royal Society Interface **12**: 20141010.

Mayr, E., 1954 Geographic speciation in tropical echinoids. Evolution **8**: 1–18.

Mayr, E., 1970 18, pp. 347–350 in *Populations, Species, and Evolution*, Harvard University Press, Cambridge, Mass.

Muller, H., 1942 Isolating mechanisms, evolution and temperature. Biol. Symp. **6**: 71–125.

Nee, S., 2001 Inferring speciation rates from phylogenies. Evolution **55**: 661–668.

Neher, R. A. and B. I. Shraiman, 2011 Statistical genetics and evolution of quantitative traits. Rev. Mod. Phys. **83**: 1283–1300.

Orr, H. A., 1995 The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics **139**: 1805–1813.

Orr, H. A. and M. Turelli, 2001 The evolution of postzygotic isolation: accumulating dobzhansky-muller incompatibilities. Evolution **55**: 1085–1094.

Owen, R., R. Crossley, T. Johnson, D. Tweddle, I. Kornfield, S. Davison, D. Eccles, and D. Engstrom, 1990 Major low levels of lake malawi and their implications for speciation rates in cichlid fishes. Proceedings of the Royal Society of London B: Biological Sciences **240**: 519–553.

Rubinoff, R. W. and I. Rubinoff, 1971 Geographic and reproductive isolation in atlantic and pacific populations of panamanian bathygobius. Evolution pp. 88–97.

Santos, M. E. and W. Salzburger, 2012 Evolution. how cichlids diversify. Science **338**: 619–621.

Schaper, S. and A. A. Louis, 2014 The arrival of the frequent: how bias in genotype-phenotype maps can steer populations to local optima. PloS one **9**: e86635.

Shea, M. A. and G. K. Ackers, 1985 The or control system of bacteriophage lambda. a physical-chemical model for gene regulation. J Mol Biol **181**: 211–230.

Spitz, F. and E. E. Furlong, 2012 Transcription factors: from enhancer binding to developmental control. Nature Reviews Genetics **13**: 613–626.

Ting, C.-T., S.-C. Tsaur, M.-L. Wu, and C.-I. Wu, 1998 A rapidly evolving homeobox at the site of a hybrid sterility gene. Science **282**: 1501–1504.

True, J. R. and E. S. Haag, 2001 Developmental system drift and flexibility in evolutionary trajectories. Evolution & development **3**: 109–119.

Tulchinsky, A. Y., N. A. Johnson, and A. H. Porter, 2014a Hybrid incompatibility despite pleiotropic constraint in a sequence-based bioenergetic model of transcription factor binding. Genetics **198**: 1645–1654.

Tulchinsky, A. Y., N. A. Johnson, W. B. Watt, and A. H. Porter, 2014b Hybrid incompatibility arises in a sequence-based bioenergetic model of transcription factor binding. Genetics **198**: 1155–1166.

Weinberger, E., 1991 Fourier and taylor series on fitness landscapes. Biological Cybernetics **65**: 321–330, 10.1007/BF00216965.

Weinreich, D. M., Y. Lan, C. S. Wylie, and R. B. Heckendorn, 2013 Should evolutionary geneticists worry about higher-order epistasis? Current opinion in genetics & development **23**: 700–707.

Wittkopp, P. J., B. K. Haerum, and A. G. Clark, 2008 Regulatory changes underlying expression differences within and between drosophila species. Nature genetics **40**: 346–350.

Wolf, J. B., J. Lindell, and N. Backström, 2010 Speciation genetics: current status and evolving approaches.

Wray, G. A., 2007 The evolutionary significance of cis-regulatory mutations. Nature Reviews Genetics **8**: 206–216.