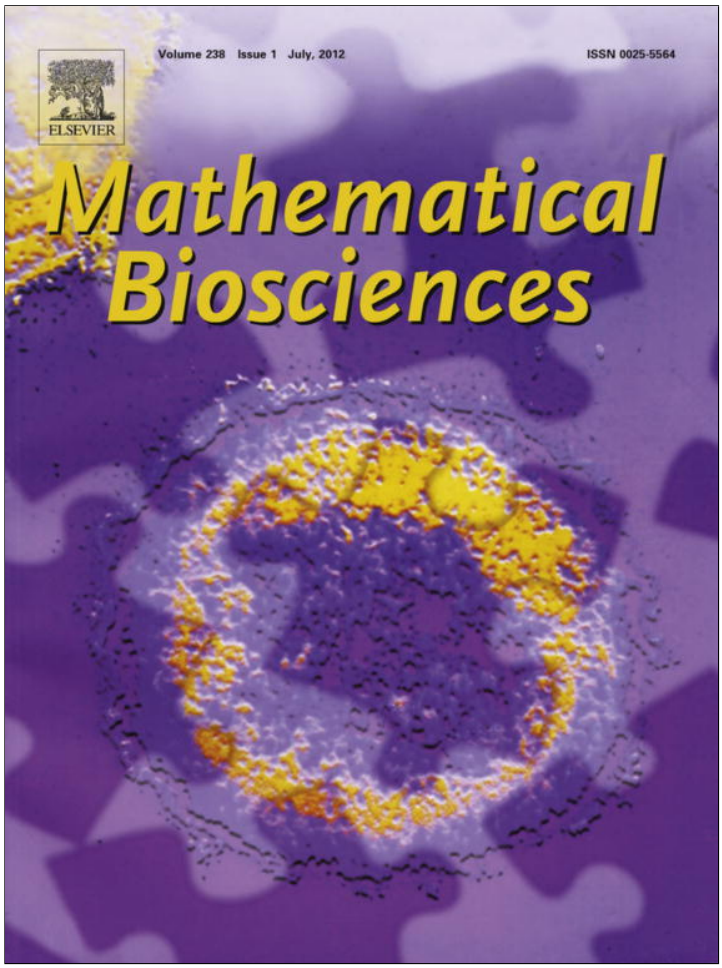

Mathematical Biosciences 238 (2012) 49–53

Contents lists available at SciVerse ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

A stochastic model for the development of Bateson–Dobzhansky–Muller incompatibilities that incorporates protein interaction networks

Kevin Livingstone [a,*], Peter Olofsson [b], Garner Cochran [b,1], Andrius Dagilis [a,1], Karen MacPherson [a,1], Kerry A. Seitz Jr [a,1]

[a] Department of Biology, Trinity University, 1 Trinity Place, San Antonio, TX 78212, United States
[b] Department of Mathematics, Trinity University, 1 Trinity Place, San Antonio, TX 78212, United States

ARTICLE INFO

Article history:
Received 1 November 2011
Received in revised form 12 March 2012
Accepted 14 March 2012
Available online 29 March 2012

Keywords:
Bateson–Dobzhansky–Muller interactions
Protein–protein interaction networks
Reproductive incompatibility
Speciation

ABSTRACT

Speciation is characterized by the development of reproductive isolating barriers between diverging groups. Intrinsic post-zygotic barriers of the type envisioned by Bateson, Dobzhansky, and Muller are deleterious epistatic interactions among loci that reduce hybrid fitness, leading to reproductive isolation. The first formal population genetic model of the development of these barriers was published by Orr in 1995, and here we develop a more general model of this process by incorporating finite protein–protein interaction networks, which reduce the probability of deleterious interactions *in vivo*. Our model shows that the development of deleterious interactions is limited by the density of the protein–protein interaction network. We have confirmed our analytical predictions of the number of possible interactions given the number of allele substitutions by using simulations on the *Saccharomyces cerevisiae* protein–protein interaction network. These results allow us to define the rate at which deleterious interactions are expected to form, and hence the speciation rate, for any protein–protein interaction network.

1. Introduction

A long-recognized hallmark of speciation is the development of intrinsic reproductive isolating barriers (RIB). As evolutionary principles were being reconciled with modern genetics, Bateson [2], Dobzhansky [5], and Muller [16] independently derived genetic models that allowed for the development of these barriers in diverging lineages. All of these models, now collectively called the BDM model, describe how fixation of mutations at two or more loci in different populations could produce inviability or sterility in hybrid offspring, without the mutations causing lowered fitness within either population. Briefly, the BDM model starts with an ancestral population of genotype *aabb*; in one population, the *A* allele arises and becomes fixed, while in the other population, *B* arises and is fixed. The resulting hybrid from the *AAbb* × *aaBB* cross would have genotype *AaBb*, and as *A* and *B* have never been 'tested' together, they could behave epistatically to cause a deleterious incompatibility. The accumulation of such Bateson–Dobzhansky–Muller incompatibilities (BDMIs) can cause permanent isolation, and hence speciation. Recent empirical work in a variety of taxa has led to the genetic characterization of BDMIs, including cloning of the loci involved [1,21,22].

Although the BDM model for the development of RIBs was widely accepted, relatively few efforts were made to extend this theory until Orr's landmark paper [17], which has subsequently been elaborated on by many others [26,8,12,7,20,6]. In the basic Orr model, two diverging lineages fix new alleles at *K* loci between them, and each new allele that arises has a probability *p* of causing a negative interaction with any of the alleles in the other genome where a substitution has occurred. One of the main insights to come from this model is that the probability of speciation rises as a function of K^2, a phenomenon that has come to be known as the snowball effect. This snowballing of BDMIs has recently been described in both *Drosophila* [13] and *Solanum* [15].

We know, however, that not all genes in the genome interact with each other in a way that could lead to the possibility of BDMIs between them. In fact, we have learned through genomics and proteomics techniques that most proteins in a protein–protein interaction (PPI) network are connected to only a small number of other proteins, while very few proteins act as central hubs with myriad interactions [10,27]. While Orr recognized the importance of interaction networks in a later paper [19], no formal treatment of complex networks was developed. In this work, we incorporate the structure of finite PPI networks to create a more general model for the development of BDMIs.

* Corresponding author. Tel.: +1 (210) 999 7236; fax: +1 (210) 999 7229.
E-mail address: klivings@trinity.edu (K. Livingstone).
[1] Equal contributors.

0025-5564/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.mbs.2012.03.006

2. Methods

2.1. Model

The starting point for the model is a network of the interactions between loci present in the most recent common ancestor of two diverging groups. Interactions here are defined broadly, and could be physical, genetic, or biochemical, so long as there exists a way for the loci to possibly cause a BDMI. We treat interaction networks as undirected graphs where each node of the graph represents a locus, and each existing interaction is denoted by an edge. These graphs have no parallel edges, and because they are undirected, there is no distinction between edges (a, b) and (b, a). The process of developing BDMIs proceeds by randomly selecting nodes without replacement, which corresponds to mutations being fixed in either lineage, allowing only one mutation per locus. A potential BDMI arises when both nodes connected by an edge are selected, and the number of potential BDMIs after K mutations is a random variable, X_K.

We consider three different protein interaction network types that we call the complete, biological, and disjoint networks (Fig. 1). The parameters we use to describe the properties of each graph and the speciation process are N, the number of nodes in the graph; N_E, the number of edges in the graph; and K, the number of substitutions occurring in either lineage since divergence. In the complete network (Fig. 1A), every protein has an interaction with every other protein, and $X_K \equiv \binom{K}{2}$. In this formulation, speciation would occur as described in Orr's original model [17].

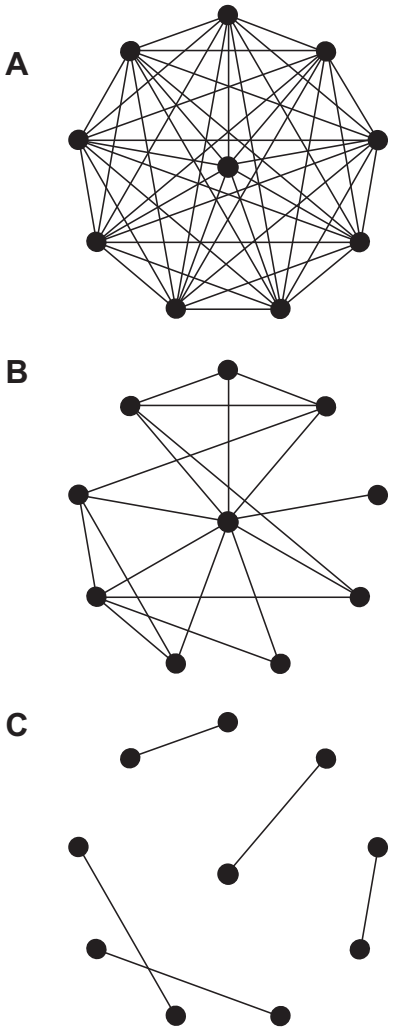

Fig. 1. Classes of graph topologies considered in this study. (A) The complete graph, where all nodes connect to all other nodes. (B) The biological graph, which models networks with high variability in connectivity among nodes. (C) The disjoint graph, where each node connects to only one other node. In each graph nodes are considered to be loci and edges represent interactions between the two loci.

The second network we consider is termed the biological network (Fig. 1B). PPI networks characterized to date follow a power law [27], which translates to order of magnitude differences in the number of edges per node in the network, which we reasoned would affect X_K. As a starting point for modeling a biological PPI network, we used the database of physical and genetic interactions from *Saccharomyces cerevisiae* in BioGrid Release 2.0.55 [25], which contains 6018 nodes and 157,861 edges after removing duplicates and converting directed edges to undirected edges. While previous studies have shown that network databases can have high false positive and negative rates [23], to our knowledge *S. cerevisiae* has the most complete and reliable of the PPI networks, and thus makes the best model.

The disjoint network (Fig. 1C) models speciation through a process of reciprocal silencing of gene duplicates, a phenomenon seen in *Arabidopsis* [3] and *Oryza sativa* [14]. In this graph, the nodes represent pairs of gene duplications present in the ancestral genome, and these pairs are connected by an edge because an organism must have at least one functioning duplicate in order to survive. Reproductive isolation in this model occurs when an ancestral population with duplicated loci *A1* and *A2* splits, and in one population the *A1* allele is silenced to give genotype $A1^sA1^s$ *A2A2*, while in the second population the other duplicate is silenced, giving the genotype *A1A1* $A2^sA2^s$. The hybrid of these two populations would be genotype $A1A1^s$ $A2A2^s$, which would be viable because it possesses a functional copy of both genes, but a proportion of selfed progeny from this individual would have genotype $A1^sA1^s$ $A2^sA2^s$, and would therefore be inviable or sterile. In this model, small islands of the genome become isolated first, which could eventually lead to complete isolation. Analytically, the disjoint graph represents a 'worst case' scenario for speciation due to the fact that it has the lowest possible number of edges $(\frac{N}{2})$ in a graph where all nodes have at least one connection.

2.2. Analytical methods

In Orr's model, the probability of speciation, S, is given by:

$$S = 1 - \prod_{n=1}^{K}(1-p)^{n-1} = 1 - (1-p)^{\binom{K}{2}} \tag{1}$$

where K is the number of mutations in both lineages, and p is the probability of a deleterious interaction between two loci. In our model, we let X_K be a random variable representing the number of potential BDMIs after K mutations, such that the conditional probability of speciation can be expressed as

$$S = 1 - (1-p)^{X_K} \tag{2}$$

which we note is itself a random variable that depends on X_K. An expression for the (unconditional) probability of speciation is obtained by computing the expected value of S:

$$E[S] = 1 - \sum_{j=1}^{N_E}(1-p)^j P(X_K = j) \tag{3}$$

Because this probability needs the distribution of X_K, which may be difficult to express, we use a first order Taylor expansion to approximate $E[S]$ as a function of the expected number of potential BDMIs, $E[X_K]$, between the K selected loci:

$$E[S] \approx 1 - (1-p)^{E[X_K]} \tag{4}$$

For any network structure, we must thus find $E[X_K]$, and a general way to do this given the variety of topologies is to use an indicator function. An indicator function I is a random variable that takes on values in $\{0,1\}$, as determined by whether some event is considered a success, in which case $I = 1$. We can enumerate the edges in the graph from 1 to N_E, and for edge j let success be defined by selecting its two nodes. Then

$$X_K = \sum_{j=1}^{N_E} I_j \tag{5}$$

and by additivity of expected values:

$$E[X_K] = \sum_{j=1}^{N_E} E[I_j] = \sum_{j=1}^{N_E} 1 \cdot P(I_j = 1) \tag{6}$$

We can compute $P(I_j = 1)$ by considering that there are $\binom{N}{K}$ possible sets of affected loci after K substitutions, and $\binom{N-2}{K-2}$ sets remaining if the two nodes of a specific edge must be included in the set of K substitutions, leading to

$$P(I_j = 1) = \frac{\binom{N-2}{K-2}}{\binom{N}{K}} = \frac{K(K-1)}{N(N-1)} \quad \text{for all } j$$

and

$$E[X_K] = N_E \frac{K(K-1)}{N(N-1)} = \alpha \cdot \binom{K}{2} \tag{7}$$

where $\alpha = \frac{N_E}{\binom{N}{2}}$, the density of the network. The probability of speciation can now be expressed as follows:

$$E[S] \approx 1 - (1-p)^{\alpha\binom{K}{2}} \tag{8}$$

Note that $E[X_K]$ is not dependent on the specific distribution of edges, only on K, N_E, and N, and so this equation can be applied to any network topology.

In order to more fully describe the trajectory of speciation, we can also use a first order Taylor expansion to provide the variance of S, for which we need $Var[X_K]$. While we can find a general function directly relating α and $E[X_K]$, the specific structure of a network impacts the variance of X_K, which can be calculated as follows:

$$Var[X_K] = N_E \cdot P_2(1-P_2) + 2 \cdot \left(N_s \cdot P_3 + \left(\binom{N_E}{2} - N_S \right) \cdot P_4 - \binom{N_E}{2} \cdot P_2^2 \right) \tag{9}$$

where N_E = the number of edges total, N_S = the number of edge pairs that share a node, $P_2 = \frac{\binom{N-2}{K-2}}{\binom{N}{K}}$, $P_3 = \frac{\binom{N-3}{K-3}}{\binom{N}{K}}$, and $P_4 = \frac{\binom{N-4}{K-4}}{\binom{N}{K}}$. For proof of this equation, see the Appendix. Since no edge pairs share a node in the disjoint model, the equation for the variance in this graph is reduced to the following:

$$Var[X_K] = N_E \cdot P_2(1-P_2) + 2 \cdot \left(\binom{N_E}{2} \cdot P_4 - \binom{N_E}{2} \cdot P_2^2 \right) \tag{10}$$

In the complete graph the variance is naturally 0.

We can use $Var[X_K]$ to calculate the variance of $S = 1 - (1-p)^{X_K}$ using the first order Taylor expansion:

$$Var[S] \approx Var[X_K] \cdot (1-p)^{2 \cdot E[X_K]} (\log(1-p))^2 \tag{11}$$

3. Results and discussion

The accuracy of the approximation in Eq. (8) to the exact expression in Eq. (3) was examined for the three networks. For the complete network, Eqs. (3) and (8) are identical because $X_K \equiv \binom{K}{2}$. For the disjoint network, the probabilities $P(X_K = j)$ can be computed explicitly as

$$P(X_K = j) = \frac{\binom{N/2}{j}\binom{N/2-j}{K-2j} 2^{K-2j}}{\binom{N}{K}} \tag{12}$$

for $j = 0, \ldots, \lfloor K/2 \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor function (for a proof, see the Appendix). Calculations reveal that Eqs. (3) and (8) agree very well over wide ranges of K, N, and p values for the disjoint network.

The agreement between Eqs. (3) and (8) in the biological network was examined by running simulations for values of K from 5 to 490, in increments of 5. In each simulation, K nodes were chosen without replacement assuming a $1/N$ probability of selecting any node, and the number of edges between these K nodes, X_K, was counted. The process was repeated for each value of K until the frequencies converged on stable estimates. The frequencies were then used to estimate the probabilities $P(X_K = j)$ in (3) to compute $E[S]$ and compare to Eq. (8). As with the other two network models, the two expressions also agree well for the biological network.

The fact that $E[X_K]$ can be computed from the network density alone leads to questions of how sensitive this analysis is to the reliability of the PPI dataset, especially given questions of high false positive and negative rates in these datasets [23]. We note that changes in N_E versus changes in α are proportional for fixed N, and consequently false positive and negative rates are less of a concern when N is large. As an example, in the case of the yeast data, adding in or taking away 10,000 edges only increases or decreases α by 6.3%.

We can use this model to examine the dynamics of speciation in each of the three networks we have considered. We have shown that the probability of speciation is dependent on the density of the network, α, and the probability of an interaction being deleterious, p. As stated before, the complete network is the finite representation of Orr's model, with $\alpha = 1$. In a complete network comparable in size to the yeast PPI network ($N = 6018$), incompatibilities snowball and speciation proceeds at a very rapid rate for $10^{-5} < p < 10^{-2}$ (Fig. 2A). In the slowest scenario with $p = 10^{-5}$, the cumulative probability of speciation is 1 after substitutions have occurred at 20% of the loci.

The disjoint model represents the need for at least one functional member from a pair of gene duplicates in the genome. We have modeled speciation in a disjoint network with 6018 loci, representing a genome comprised entirely of 3009 duplicated loci (Fig. 2C); while this is clearly implausible biologically, it does allow for comparisons to the other networks. The disjoint network has $\alpha = \frac{1}{N-1}$, the lowest possible value for α in a network where each node has at least one edge, leading to the slowest predicted rate of speciation. Interestingly, if we assume that intraspecific variation reflects recent fixation, this prediction is at odds with observations of BDMI caused by reciprocal silencing of duplicates among populations of *Arabidopsis* [3] and *Oryza* [14]. These results could be reconciled by considering that mutations affecting duplicates occur more frequently, and that these mutations could have $p = 1$.

Speciation in the biological network proceeds at rates intermediate to the complete and disjoint networks (Fig. 2B). As the biological network presumably best approximates the real dynamics of BDMI accumulation, we can use it as a null model to look at speciation dynamics and to refine parameter estimates given additional assumptions and data. As an example, we can use this model to examine p. True biological estimates for p would be extremely difficult to obtain, as multiple alleles would need to be created for both loci in an interaction, with all pairwise combinations evaluated to determine the phenotypic effects. As this would be possible but painstaking in only a few organisms, models such as this offer us one way to gain insight into this parameter.

We can use this model to obtain a lower bound for p by following Orr in defining the random variable K_S, the mutation that causes reproductive isolation. In our model, the probability that $K_S > N$ is non-zero, which makes K_S not well defined. It seems

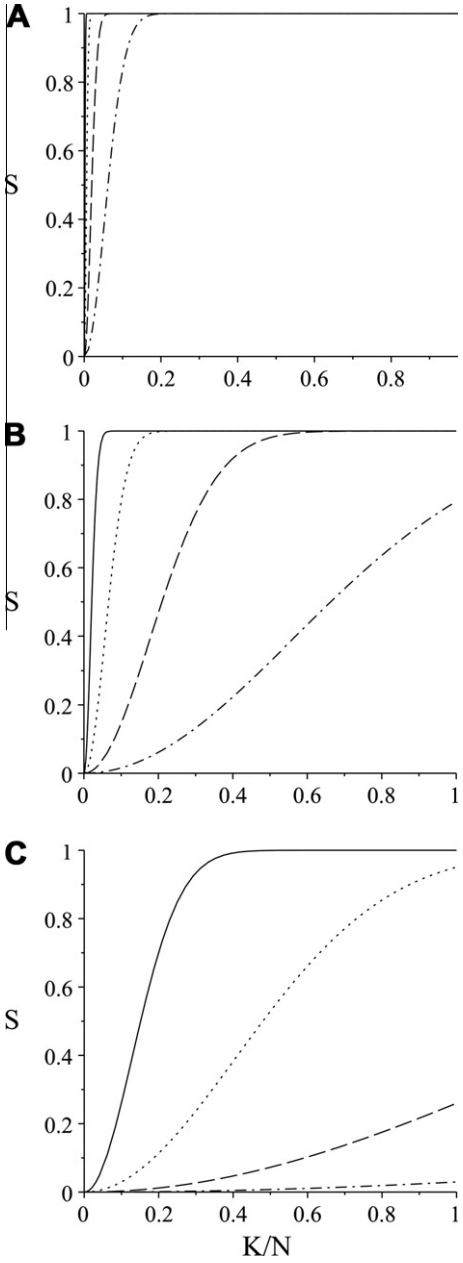

Fig. 2. Cumulative speciation probability curves vs. fraction of substituted loci for the three different networks. Curves for the (A) complete, (B) biological, and (C) disjoint networks are shown. In all plots, K is the number of substitutions, $N = 6018$ and $p = 10^{-2}$ (solid), 10^{-3} (dots), 10^{-4} (dashed), or 10^{-5} (dash-dot).

reasonable, however, to assume that speciation will occur before all loci in the genome have undergone substitutions, which makes $P(K_S > N) = (1-p)^{N_E}$ negligible. In effect, a reasonable range for p can be bounded by N_E alone. In the case of the yeast PPI network, if we set $P(K_S > N) \leqslant 0.001$, we estimate that $p \geqslant 4.4 \times 10^{-5}$.

We can also define the high end of the range for p by combining this model with experimental data. Under the assumption that speciation occurs before all loci in the genome have undergone substitution, we can compute $E[K_S]$ conditional on $K_S \leqslant N$:

$$E[K_S | K_S \leqslant N] \approx \sqrt{\frac{\pi}{2\alpha p}} \tag{13}$$

(for derivation see [17]). We can then apply this equation to the example of laboratory speciation in *S. cerevisiae*, where a BDMI was seen after 500 generations of strong divergent selection and 17 confirmed allelic substitutions [1]. Substituting 17 for K_S gives an estimate of $p \approx 0.6$ in this system, which admittedly seems rather high. It is possible the high value for p estimated from the *S. cerevisiae* data could be an artifact of the particular loci under selection in these divergent environments. This hypothesis could be tested by repeated lab trials under a variety of paired environments, which could give both true mean values for K_S for each condition and means for a variety of conditions, providing an upper bound for p. We recognize that these two estimates for p vary over four orders of magnitude, and that the answer will most likely lie somewhere in the middle.

Another important use of this model is that it forms a framework for considering BDMIs that arise through more complex interactions. It has been shown that the number of possible complex interactions rises dramatically more quickly than the number of pairwise interactions [28], and that complex interactions have been seen to cause RIBs in *Saccharomyces sensu stricto* yeasts [11] and *Drosophila* [4,18]. Some well established models look at such interactions [8,9], but it is difficult to make progress with these models primarily because there is no good definition for what constitutes a complex interaction. Using this model we can define complex interactions as subgraphs in the network with specific levels of connectivity, as paths between loci, or any other biologically plausible way that multiple loci could interact. These definitions could lead to either analytical or simulation-based results that shed light on the question of the relative frequency of complex vs. simple interactions in the speciation process. One thing we can be certain of, however, is that incomplete biological networks will slow this process relative to complete networks, and our model does provide a framework for such investigations.

Finally, we note that our model presents an intriguing hypothesis as to why taxonomic groups may have correlated rates of speciation [24]. It can be assumed that the protein interaction networks for related taxa are of a similar size and density, which could lead to similar speciation rates under this model. While other factors undoubtedly play a role in these processes, similarities in PPI networks among related taxa could account for at least part of this phenomenon.

Acknowledgments

We gratefully acknowledge financial support from the Howard Hughes Medical Institute through a grant to Trinity University and the NSF through grant UBM-0926702.

Appendix A. Proofs

Proof of Eq. (9). In order to find the variance an approach must be made which takes into account the structure of individual nodes. Recall that

$$X_K = \sum_{j=1}^{N_E} I_j$$

and hence

$$Var[X_K] = \sum_{j=1}^{N_E} Var[I_j] + \sum_{i \neq j=1}^{N_E} Cov[I_i, I_j]$$

where

$$Var[I_j] = E[I_j^2] - E[I_j]^2$$

and

$$Cov[I_i, I_j] = E[I_i \cdot I_j] - E[I_i]E[I_j]$$

Since $I_j \in \{0, 1\}$ it follows that

$$E[I_j^2] = E[I_j] = P(I_j = 1) = P_2$$

so

$$Var[I_j] = P_2(1 - P_2)$$

For the covariance,

$$E[I_i] = E[I_j] = P_2$$

and

$$E[I_i \cdot I_j] = P(I_i = 1 \quad \text{and} \quad I_j = 1)$$

the latter probability being dependent on whether the pair of edges share a node, or if they are disjoint. If they share a node, then

$$P(I_i = 1 \quad \text{and} \quad I_j = 1) = \frac{\binom{N-3}{K-3}}{\binom{N}{K}} = P_3$$

If they do not share a node, then

$$P(I_i = 1 \quad \text{and} \quad I_j = 1) = \frac{\binom{N-4}{K-4}}{\binom{N}{K}} = P_4$$

Using the notation from the expected value,

$$\sum_{i \neq j=1}^{N_E} Cov[I_i, I_j] = 2 \cdot \sum_{i<j=1}^{N_E} Cov[I_i, I_j]$$
$$= 2 \cdot \left((N_S \cdot P_3 + N_D \cdot P_4) - \binom{N_E}{2} \cdot P_2^2 \right)$$

where N_S is the number of edge pairs which share a node, and N_D is the number of edge pairs that do not share a node. Therefore

$$Var[X_K] = N_E \cdot P_2(1 - P_2) + 2 \cdot \left(N_s \cdot P_3 + \left(\binom{N_E}{2} - N_S \right) \cdot P_4 - \binom{N_E}{2} \cdot P_2^2 \right)$$

Proof of Eq. (12). In the disjoint network, there are N nodes and $N/2$ edges. We choose K nodes at random and consider the event that we get $X_K = j$ edges. The total number of choices equals $\binom{N}{K}$ and to get the expression in the numerator, first note that there are $\binom{N/2}{j}$ sets of j edges. To get a specific such set, we need to choose the $2j$ nodes in it and the remaining $K - 2j$ nodes must be chosen among the remaining $N/2 - j$ nodes such that no more edges are included. In other words, the remaining nodes must be chosen by choosing one of the remaining $N/2 - j$ edges for each of the $K - 2j$ nodes, and there are $\binom{N/2-j}{k-2j}$ such choices. Finally, since each edge has 2 nodes, there are 2^{K-2j} possible node arrangements for a given set of $K - 2j$ edges and Eq. (12) follows.

References

[1] J.B. Anderson, J. Funt, D.A. Thompson, S. Prabhu, A. Socha, C. Sirjusingh, J.R. Dettman, L. Parreiras, D.S. Guttman, A. Regev, L.M. Kohn, Determinants of divergent adaptation and Dobzhansky–Muller interaction in experimental yeast populations, Current Biology 20 (15) (2010) 1383.

[2] W. Bateson, Heredity and variation in modern lights, in: A.C. Seward (Ed.), Darwin and Modern Science, Cambridge University Press, 1909, p. 85.

[3] D. Bikard, D. Patel, C. Le Metté, V. Giorgi, C. Camilleri, M.J. Bennett, O. Loudet, Divergent evolution of duplicate genes leads to genetic incompatibilities within A. *thaliana*, Science 323 (5914) (2009) 623.

[4] E.L. Cabot, A.W. Davis, N.A. Johnson, C.-I. Wu, Genetics of reproductive isolation in the drosophila simulans clade: complex epistasis underlying hybrid male sterility, Genetics 137 (1994) 175.

[5] T. Dobzhansky, Genetic nature of species differences, American Naturalist 71 (735) (1937) 404.

[6] J.L. Fierst, T.F. Hansen, Genetic architecture and postzygotic reproductive isolation: Evolution of Bateson–Dobzhansky–Muller incompatibilities in a polygenic model, Evolution 64 (3) (2010) 675.

[7] B.M. Fitzpatrick, Hybrid dysfunction: population genetic and quantitative genetic perspectives, American Naturalist 171 (4) (2008) 491.

[8] S. Gavrilets, Perspective: models of speciation: what have we learned in 40 years?, Evolution 57 (10) (2003) 2197

[9] S. Gavrilets, Fitness landscapes and the origin of species Sergey Gavrilets, Monographs in population biology, 41, Princeton University Press, Princeton, NJ, 2004.

[10] H. Jeong, S.P. Mason, A.L. Barabási, Z.N. Oltvai, Lethality and centrality in protein networks, Nature 411 (6833) (2001) 41.

[11] K.C. Kao, K. Schwartz, G. Sherlock, A genome-wide analysis reveals no nuclear Dobzhansky–Muller pairs of determinants of speciation between S. cerevisiae S. *paradoxus* but suggests more complex incompatibilities, PLoS Genetics 6 (7) (2010) e1001038.

[12] A.S. Kondrashov, Accumulation of Dobzhansky–Muller incompatibilities within a spatially structured population, Evolution 57 (1) (2003) 151.

[13] D.R. Matute, I.a. Butler, D.a. Turissini, J.a. Coyne, A test of the snowball theory for the rate of evolution of hybrid incompatibilities, Science 329 (5998) (2010) 1518.

[14] Y. Mizuta, Y. Harushima, N. Kurata, Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes, Proceedings of the National Academy of Sciences USA 107 (47) (2010) 20417.

[15] L.C. Moyle, T. Nakazato, Hybrid incompatibility snowballs between Solanum species, Science 329 (5998) (2010) 1521.

[16] H.J. Muller, Isolating mechanisms, evolution and temperature, Biological Symposia 6 (1942) 71–125.

[17] H.A. Orr, The genetics of speciation: the evolution of hybrid incompatibilities, Genetics 139 (1) (1995) 1805.

[18] H.a. Orr, S. Irving, Complex epistasis and the genetic basis of hybrid sterility in the Drosophila pseudoobscura Bogota-USA Hybridization, Genetics 158 (3) (2001) 1089.

[19] H.a. Orr, M. Turelli, The evolution of Postzygotic isolation: accumulating Dobzhansky–Muller incompatibilities, Evolution 55 (6) (2001) 1085.

[20] M.E. Palmer, M.W. Feldman, Dynamics of hybrid incompatibility in gene networks in a constant environment, Evolution 63 (2) (2009) 418.

[21] D.C. Presgraves, The molecular evolutionary basis of species formation, Nature Reviews Genetics 11 (3) (2010) 175.

[22] L.H. Rieseberg, B.K. Blackman, Speciation genes in plants, Annals of Botany 106 (3) (2010) 439.

[23] A.W. Rives, T. Galitski, Modular organization of cellular networks, Proceedings of the National Academy of Sciences USA 100 (3) (2003) 1128.

[24] J.J. Sepkoski, Rates of speciation in the fossil record, Philosophical Transactions of the Royal Society of London Series B Biological Sciences 353 (1366) (1998) 315.

[25] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, M. Tyers, BioGRID: a general repository for interaction datasets, Nucleic Acids Research 34 (Database issue) (2006) D535.

[26] M. Turelli, H.a. Orr, Dominance,epistasis and the genetics of postzygotic isolation, Genetics 154 (4) (2000) 1663.

[27] A. Wagner, The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes, Molecular Biology and Evolution 18 (7) (2001) 1283.

[28] J.J. Welch, Accumulating Dobzhansky–Muller incompatibilities: reconciling theory and data, Evolution 58 (6) (2004) 1145.