Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 242 (2006) 643–651
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Competitive signal discrimination, methylation reprogramming and genomic imprinting

Jon F. Wilkins*

Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Received 1 November 2005; received in revised form 4 April 2006; accepted 26 April 2006
Available online 9 May 2006

Abstract

Genomic imprinting (parent-of-origin-dependent gene regulation) is associated with *intra*-genomic evolutionary conflict over the optimal pattern of gene expression. Most theoretical models of imprinting focus on the conflict between the maternally and paternally derived alleles at an imprinted locus. Recently, however, more attention has been focused on multi-directional conflicts involving not only the imprinted gene itself, but also the genes that encode the regulatory machinery responsible for establishing and maintaining imprinted gene expression. In this paper, I examine the conflict involved in epigenetic reprogramming of imprinted genes in early mammalian embryonic development. In the earliest phase of development, maternal-store proteins are responsible for most regulatory activity in the embryo. These proteins are under selection to maximize the mother's inclusive fitness, which is not identical to that of either of the sets of genes present in the embryo. Both the maternally and paternally derived genomes in the embryo favor maintenance of the epigenetic modifications established in the female and male germlines, respectively. Maternal-store proteins favor maintenance of some of these modifications, but erasure of others. Here I consider the logical structure of the machinery responsible for these two activities. Methylation maintenance is most effectively performed by *AND*-linked architectures, which may explain the unusual trafficking behavior of the oocyte-specific DNA methyltransferase, Dnmt1o. By contrast, demethylation is better supported by *OR*-linked architectures, which may explain the difficulty in identifying the factor(s) responsible for the active demethylation of the paternal genome following fertilization.

Keywords: Genomic imprinting; Dnmt1o; Signal discrimination; Epigenetic reprogramming; Methylation

1. Introduction

Genomic imprinting is the phenomenon where a gene's pattern of expression depends on whether it was inherited from a male or a female in the previous generation (Delaval and Feil, 2004; Jaenisch and Bird, 2003; Morison et al., 2005). According to the Kinship Theory of Imprinting, this is the outcome of an evolutionary conflict between the maternally and paternally derived alleles within an organism (Haig, 2000; Haig and Westoby, 1989; Wilkins and Haig, 2003). Simple evolutionary models have been successful in explaining many of the most obvious aspects of imprinting (Haig, 2004). In particular, for genes that are expressed in fetal or placental tissues in mammals and that influence the fetal growth rate, an imprinted growth enhancer is expected to be transcriptionally silent on the maternally derived copy; conversely, imprinted growth suppressors are paternally silenced.

Under these models, monoallelic expression is an evolutionarily stable strategy (ESS) that evolves in response to a conflict over the optimal level of gene expression (Haig, 1997; Mochizuki et al., 1996; Wilkins and Haig, 2001). One limitation of these simple models is that they assume (often implicitly) that the expression of an allele at an imprinted locus is under sole control of the allele itself—or rather, under the control if *cis*-acting regulatory sequences that are tightly linked to the imprinted gene. However, the establishment, maintenance, and interpretation of the epigenetic marks associated with imprinting involve interactions between *cis*-acting and *trans*-acting elements. In most cases, these *trans*-acting elements represent "genetic factions" whose evolutionary interests are not identical to either the maternally derived

*Tel.: +1 505 946 2755; fax: +1 505 982 0565.
E-mail address: wilkins@santafe.edu.

0022-5193/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2006.04.015

or paternally derived allele at an imprinted locus. Recent theoretical work has sought to extend the logic of the Kinship theory to these other genetic factions.

In particular, a number of recent studies have focused on the distinction between *imprinted* genes, which exhibit parent-of-origin-specific gene expression, and *imprinting* genes, whose gene products participate in the establishment of imprinting (Burt and Trivers, 1998; Spencer and Williams, 1997; Wilkins and Haig, 2002). For example, the (*trans*-acting) DNA methyltransferases expressed in the male and female germ lines are subject to evolutionary forces different from those acting on the (*cis*-acting) regulatory sequences that they modify.

In this paper, I focus specifically on the evolutionary conflicts that arise in the earliest phase of embryogenesis, prior to transcriptional activation of the embryo's genes. Immediately following fertilization, the newly formed zygotic genome undergoes a large-scale epigenetic reprogramming. This process includes modifications to chromatin-associated proteins, including histone methylation and acetylation. The best understood epigenetic modification is DNA methylation of cytosine residues at CpG dinucleotides. Reprogramming includes active demethylation of some loci and de novo establishment of methylation at others prior to the first round of cell division. Further reprogramming occurs during the first few rounds of cell division, including passive demethylation (failure to propagate methylation in conjunction with DNA replication).

These epigenetic processes are locus-specific (Bestor, 2003; Gaudet et al., 2004; Lane et al., 2003; Rakyan et al., 2003; Santos et al., 2002), and vary among mammalian species. This indicates that the machinery involved in the removal, propagation, and spreading of DNA methylation discriminates among classes of DNA sequences, and that this discriminatory mechanism may be subject to natural selection.

In the earliest stages of embryogenesis, many cellular processes are performed by maternal-store proteins—gene products representing both of the mother's alleles. While the phenotypic effects of these genes manifest in the offspring, natural selection acts on their consequences for the inclusive fitness of the mother. Maternal-store proteins that modify the epigenetic state of genes in the embryo will do so in the interests of the mother, even if those interests are not identical to those of the offspring, or of sets of the offspring's alleles (Burt and Trivers, 1998; Moore and Reik, 1996; Wilkins and Haig, 2002). Specifically, many imprinted genes affect the distribution of maternal resources among offspring, and maternal genes will favor modifications that reduce the demand that individual offspring place on the mother (Haig, 1993, 1996).

2. Methylation reprogramming and competitive signal discrimination

More generally, maternal genes tend to benefit from the effects of methylation established in oogenesis, but not spermatogenesis, and will favor eliminating methylation specifically from paternally derived chromosomes (Reik and Walter, 2001; Wilkins, 2005; Wilkins and Haig, 2002). On the other hand, *cis*-acting elements associated with imprinted genes on both the maternally and paternally derived chromosomes will be under selection to preserve their own methylation patterns (Wilkins, 2005; Wilkins and Haig, 2002). Thus both demethylation and methylation maintenance are potentially undergoing antagonistic coevolution.

In this paper, I consider the coevolution of the maternally encoded epigenetic machinery the maternally and paternally derived alleles that the machinery acts upon. I consider the system more abstractly as a problem of competitive signal discrimination (CSD). That is, I treat the epigenetic machinery as a signal receiver that must distinguish between two classes of signals, when those signals are, themselves, subject to natural selection. The success of a particular receiver in performing a CSD task (and the fitness of the genes encoding that receiver) will be influenced by numerous factors, such as the nature of selection acting on the signals, specificity of signal recognition, and the rates at which signals and receivers are evolving. Here I restrict analysis to the consequences of different logical structures on the coevolutionary performance of receivers. Specifically, I consider compound *receiver architectures*, which consist of multiple *receiver systems* and logical relations among those systems. For example, imagine that an organism produces two separate receivers designed to detect a particular change in the environment, such as elevated temperature. If the two receiver systems are *AND-linked*, the organism initiates its high-temperature response only if *both* systems detect a temperature increase. By contrast, if the two systems are *OR-linked*, the high-temperature response is triggered by activation of either of the two systems.

I will begin by considering in general terms the factors that might favor different logical structures. The intuitions developed here are illustrated through the use of coevolutionary simulations. Following this general analysis, I return to the specific problem of epigenetic reprogramming of imprinted genes in the mammalian embryo.

3. Definitions

A *receiver system* ρ is defined by a set of properties that a signal S must possess in order for ρ to be activated by S. A *receiver architecture* R consists of one or more recognition systems ρ_i and logical relations among them that determine what combination of the ρ_i must be activated in order for R to perform action A. I assume that R is under selection to perform A specifically in response to signals of type S_1, but not signals of type S_2 (Fig. 1). I consider two types of selection on genes responsible for producing the signals: *evasive selection*, which favors signals that avoid recognition by R, and *associative selection*, which favors those signals that activate R (Fig. 2).

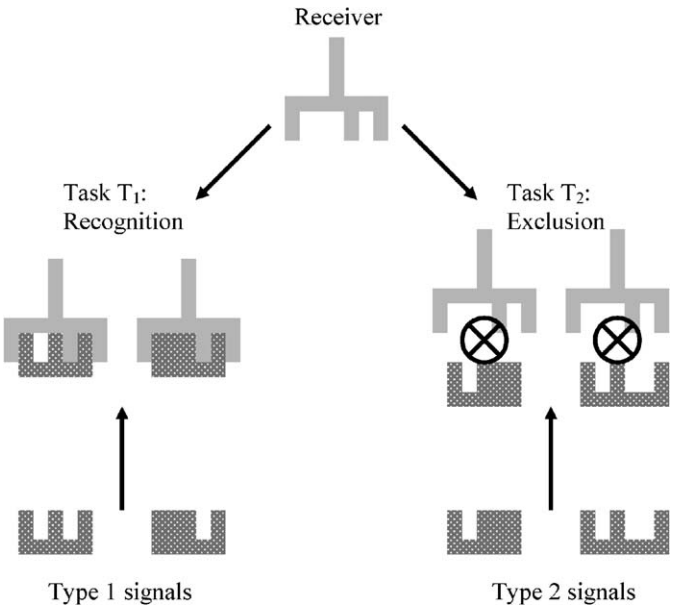

Fig. 1. The components of competitive signal discrimination. A receiver faced with a task of competitive signal discrimination must accomplish two tasks. It must activate in response to one class of signals (S_1), but fail to activate in response to a second class (S_2). These two tasks are represented here as a lock-and-key type of interaction. The receiver pictured here has successfully found an arrangement that will bind to the S_1 signals, but not to the S_2 signals.

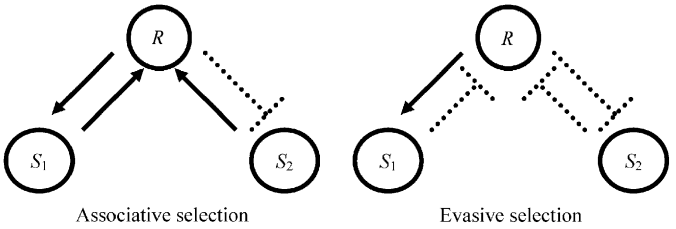

Fig. 2. Evasive vs. associative selection on signals. The two types of selection on signals considered in the text are represented schematically here. In both cases, selection on the receiver system R favors recognition of S_1 (represented by a solid arrow), and exclusion of S_2 (represented by a dashed line). Under associative selection, both S_1 and S_2 are under selection to interact with (be recognized by) R. Under evasive selection, both classes of signal are under selection to avoid recognition by R.

The CSD problem can be separated conceptually into two tasks. First, in task T_1, R must respond to S_1 by initiating action A. Second, in task T_2, R must fail to respond to S_2. The issue of logical structure can be illustrated with two simple two-system architectures: R_{AND} ($=\rho_1\ AND\ \rho_2$), which is activated by a signal only if both of its constituent systems recognize that signal, and R_{OR} ($=\rho_1\ OR\ \rho_2$), which can be activated by either one of its systems (Fig. 3).

AND-linked recognition architecture

	Receiver system ρ_1 activated	Receiver system ρ_1 not activated
Receiver system ρ_2 activated	Receiver architecture R_{AND} activated	Receiver architecture R_{AND} **NOT** activated
Receiver system ρ_2 not activated	Receiver architecture R_{AND} **NOT** activated	Receiver architecture R_{AND} **NOT** activated

OR-linked recognition architecture

	Receiver system ρ_1 activated	Receiver system ρ_1 not activated
Receiver system ρ_2 activated	Receiver architecture R_{OR} activated	Receiver architecture R_{OR} activated
Receiver system ρ_2 not activated	Receiver architecture R_{OR} activated	Receiver architecture R_{OR} **NOT** activated

Fig. 3. *AND*-linked vs. *OR*-linked signal recognition. The simulation results presented here focus on a comparison between two receiver architectures (R_{AND} and R_{OR}), each of which consists of two receiver systems (ρ_1 and ρ_2). This figure graphically indicates the rules of activation for the two architectures. In the *AND*-linked case, the receiver architecture R_{AND} is activated only if both of its component receiver systems are activated. In the *OR*-linked case, the receiver architecture R_{OR} is activated if either one (or both) of its component receiver systems are activated.

4. Receiver architectures: *OR* vs. *AND*

Selection on R will be influenced by multiple factors, including the relative difficulties of tasks T_1 and T_2, and the relative costs of failing to accomplish them. The relative difficulties of T_1 and T_2 will depend on the strength of selection acting on S_1 and S_2, and on constraints on the evolution of the signals and receivers. For instance, with

biological signals, there may be constraints on signal evolution related to other, non-signaling functions of the molecules. Constraints on molecular signals might include native folding and enzyme activity of proteins or coding by DNA sequences.

A detailed analysis of selection dynamics in a particular setting will depend on various specific details, many of which are unlikely to be known with any certainty. However, an understanding, in general terms, of the competencies associated with different logical structures may be useful in understanding certain empirical observations. Intuitively, all else being equal, R_{OR} will be more permissive than R_{AND} with respect to the range of signals that it recognizes. Thus an *OR*-linked architecture will be more naturally suited to accomplishing task T_1, which requires R to recognize S_1. Conversely, the greater stringency associated with an *AND*-linked architecture will facilitate task T_2, which requires the exclusion of S_2 from recognition.

This suggests that features of a CSD problem that complicate T_1 (or simplify T_2) should favor R_{OR}, whereas features that complicate T_2 (or simplify T_1) should favor R_{AND}. I have defined two types of selection acting on the signals: evasive and associative. Evasive selection, in which signals are evolving to evade recognition, will facilitate exclusion (T_2), but make recognition (T_1) difficult, favoring R_{OR}. Associative selection, in which signals favor recognition, will make architectures of the R_{AND} type more effective. Of course, these effects may be confounded by other features. For example, whereas mild constraints on the evolution of the signals will facilitate both T_1 and T_2, simultaneous recognition of multiple severely constrained signals might favor R_{OR}.

5. Simulations

These principles are illustrated in coevolutionary simulations. Each signal (S) and receiver system (ρ) is modeled as a string of 20 integers, each of which can take on one of five values (0–4, inclusive). A population of receiver architectures (R) was coevolved with 10 populations of signals, five of type S_1, and five of type S_2 ($N = 1000$ for each population). Simulations ran for 150,000 generations, and the mean fitness of the R population was sampled every 10 generations. Activation of each ρ_i was determined by the number of matches between the strings representing the receiver system and signal, with m being the number required for half activation. Molecular constraints were modeled by fixing f sites in each signal string. The next generation was derived by sampling with replacement, weighted by fitness, and then subjected to mutation. The simulations and selection scheme are described in detail in Appendix A.

The results of two sets of simulations are presented in Fig. 4. Figs. 5 and 6 present the mean R fitness for generations 100,000–150,000, averaged over 10 independent runs for each architecture. Fig. 5 presents the unconstrained case ($f = 0$) for a range of stringencies ($0 \leqslant m \leqslant 10$). In Fig. 6, m is fixed, and f is varied. The importance of signal evolution in shaping the receiver architecture is evident in these graphs. *OR*-linkage provides a dynamic flexibility that facilitates recognition of signals that are under selection to evade that recognition. *AND*-linkage provides additional stringency to facilitate exclusion of signals that are under selection to activate the receiver.

6. Active demethylation of the paternal genome

In mice, the paternally derived genome is subject to a large-scale active demethylation process following fertilization, but prior to the first cell division (Dean et al., 2001; Mayer et al., 2000; Oswald et al., 2000; Santos et al., 2002). It has been suggested that this represents an effort by the maternal genes to reduce the paternal influence on the embryo (Reik and Walter, 2001). In the terminology used here, both maternally and paternally derived alleles at an imprinted locus will be subject to evasive selection with respect to the active demethylation process. This suggests that natural selection might favor an *OR*-linked recognition architecture for demethylation. That is, there may have been selection for the production of multiple demethylases with partially overlapping specificities.

At present, the identity of the factor or factors responsible for the active demethylation of the paternally derived genome is unknown. Methyl-binding domain protein-2 (MBD2) was proposed as a candidate for this function (Bhattacharya et al., 1999) on the basis of in vitro results. However, this finding has been called into question,

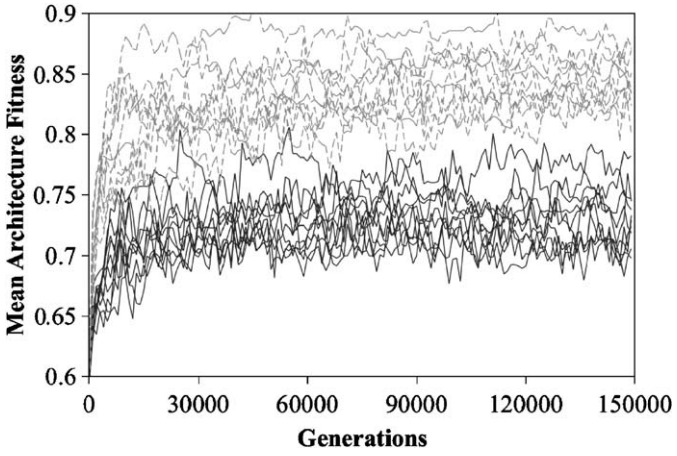

Fig. 4. Mean fitness for the recognition architecture for two sample simulation sets. The mean fitness of the population of recognition architectures was sampled once every 1000 generation over the 150,000-generation simulation run. For these simulations, $m = 6$ and $f = 10$. The population of signals was evolving under evasive selection. The solid lines represent 10 runs of the R_{AND} architecture. The dashed lines represent the same ten sets of constraints evolved against R_{OR}. R_{OR} outperforms R_{AND} under these conditions, consistent with the task of signal recognition being more difficult than signal exclusion, as expected for signals undergoing evasive selection.

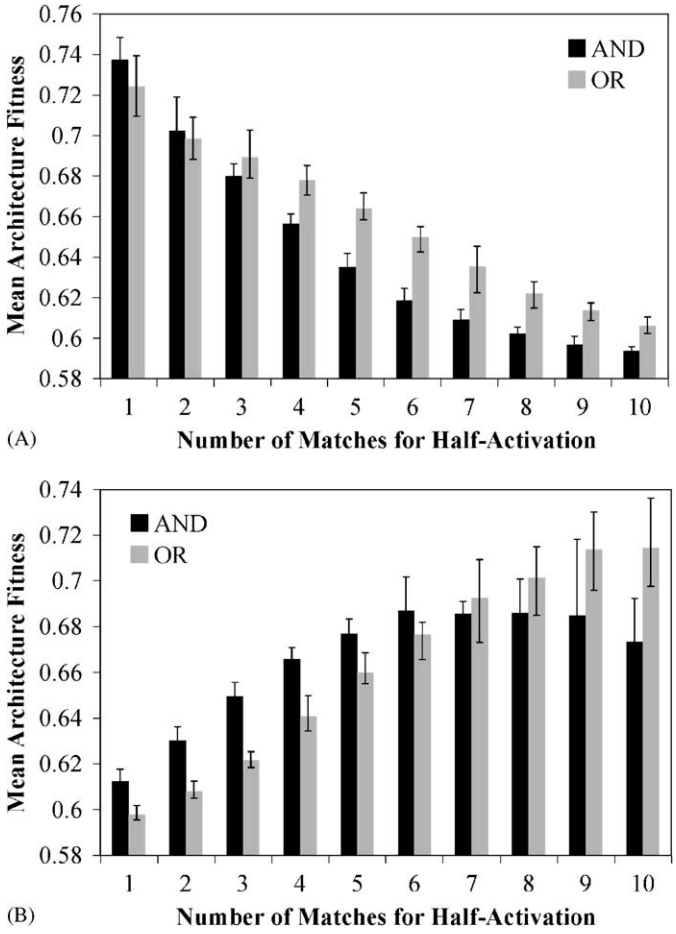

Fig. 5. Mean architecture fitness as a function of recognition stringency. These graphs compare the R_{AND} and R_{OR} architectures for a variety of values of m (the number of matches required between signal and receiver resulting in half-activation of the receiver). Increasing the stringency of recognition (increasing m) facilitates exclusion (task T_2), and therefore favors R_{OR}, whereas reducing the stringency facilitates recognition (T_1) and favors R_{AND}. The range of conditions over which R_{OR} is favored is greater under evasive selection (A) than it is under associative selection (B). Histogram height represents the mean fitness for generations 100,000–150,000 averaged over 10 independent simulation runs. Error bars indicate the highest and lowest mean fitness value among the 10 runs.

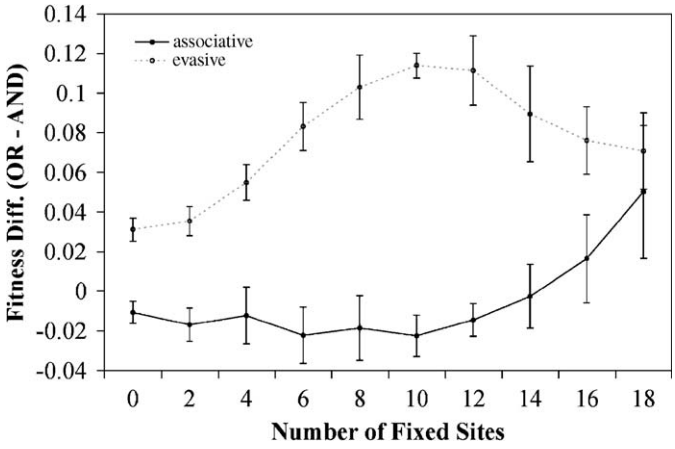

Fig. 6. R_{AND} vs. R_{OR} as a function of the severity of constraints on signal evolution. For each value of f (number of fixed sites in the signal strings, out of 20), 10 sets of constraints were randomly determined. The difference in the mean fitness between the *OR*-linked and *AND*-linked architectures was determined for each set of constraints. The average of these 10 differences is graphed here, with the error bars representing two standard deviations. The open circles represent signals under evasive selection, while the filled circles represent signals under associative selection. Evasive selection favors *OR*-linkage (positive values), while associative selection favors *AND*-linkage (negative values). Increasing the degree of constraint (increasing f) increases the difficulty of recognition, favoring *OR*-linkage. At the highest values of f, the signals are almost entirely constrained, and the form of selection acting on the unconstrained sites becomes less important, so the two curves begin to converge.

in part because other groups have had difficulty in reproducing these results (Boeke et al., 2000; Ng et al., 1999; Wade et al., 1999), and in part because disruption of the *MBD2* gene does not result in retention of paternal methylation (Santos et al., 2002). The exact role of MBD2 in epigenetic gene regulation will only be clarified through continued empirical research. However, it is worth noting that if the active demethylation architecture is, in fact, *OR*-linked, elimination of only one of the active demethylation systems might not be expected to result in a significant genome-wide retention of methylation.

More generally, as the search continues for the proteins or other factors responsible for active demethylation of the paternal genome, we should keep in mind the fact that this activity may rely on multiple, partially redundant systems. A number of mechanisms have been proposed for demethylation, involving either direct removal of the methyl group, or excision of one or more bases around the methylated site (Cedar and Verdine, 1999; Morgan et al., 2004; Trewick et al., 2002; Vairapandi, 2004; Vairapandi et al., 2000). If multiple demethylation systems do exist, they could potentially employ any combination of different mechanisms.

7. Methylation maintenance by Dnmt1o

In most somatic tissues, methylation maintenance occurs in conjunction with DNA replication, and is performed by DNA methyltransferase 1 (Dnmt1), which specifically recognizes hemimethylated DNA and methylates the cytosine in the newly synthesized strand. Methylated CpG sites that fail to recruit a maintenance methyltransferase are passively demethylated, as at each cell division the proportion of methylated cytosines is reduced by a factor of two. In mice, the somatic form Dnmt1 is absent from the preimplantation embryo. In its place is the oocyte-specific splice variant, Dnmt1o, a maternal-store protein produced during oogenesis. Through the first three rounds of cell division following fertilization, it is actively retained in the cytosol (Doherty et al., 2002). At the eight-cell stage, it moves into the nucleus, and returns to the cytosol at the 16-cell stage (Howell et al., 2001; Ratnam et al., 2002).

The trafficking behavior of Dnmt1o suggests that its participation in methylation maintenance may be specific

to the fourth round of DNA replication. Consistent with this interpretation, knockouts of Dnmt1o result in maternal-effect lethality, and a loss of methylation at imprinted loci on approximately half of the normally methylated alleles. The factor(s) responsible for maintenance at the other divisions have not yet been identified, but these results suggest that methylation maintenance in mouse embryogenesis requires at least two distinct systems.

Just as the *cis*-acting regulatory elements associated with imprinted loci will be under selection to evade recognition by the demethylation machinery, they will be under selection to facilitate recognition by the methylation maintenance machinery. In the terminology used here, they will be subject to associative selection with respect to methylation maintenance. This suggests that selection would favor an *AND*-linked recognition architecture for methylation maintenance. The apparent lack of redundancy between Dnmt1o and the one or more other maintenance methyltransferases indicates that this architecture is, in fact, *AND*-linked, consistent with the results presented above.

8. Discussion

The problem of competitive signal discrimination has many dimensions. In this paper, I have limited discussion to only two of those dimensions: associative vs. evasive selection acting on the signals, and *AND* vs. *OR* linkage among the components of the receiver architecture. I have argued that associative selection favors *AND*-linked architectures, whereas evasive selection favors *OR*-linkage. However, the effects described here do not exist in isolation. Other factors, such as recognition stringency and the severity of functional constraints, influence the relative success of different logical structures. In specific instances, these other effects may outweigh the effects that have been the focus of this work.

Furthermore, the analysis presented here is limited to a comparison of two different, preexisting architectures. I have not addressed the question of how such architectures might arise in a population-genetic context. Future analyses might consider the process by which receiver systems duplicate and diversify—or the capacity of modifiers of the logical structure to invade the population.

One aspect of the competitive signal discrimination problem that I have addressed only in passing is that of mimicry. Selection for particular types of mimicry is implicit in the coevolutionary dynamics of the model described here. Specifically, in the case of evasive selection, signals of type S_1 could benefit from mimicking S_2 signals. Similarly, when the signals are under associative selection, S_2 signals could be selected to mimic S_1. It is in this context that functional constraints are likely to be most important. The inability of individual signals to access the complete sequence space may create particular signal–receiver configurations that are resistant to mimicry.

In the specific context of the epigenetic reprogramming of imprinted genes, I have tried to indicate how these general principles might illuminate some recent empirical findings. The implication of this argument is that the logical structures associated with the active demethylation and methylation maintenance machinery have evolved in response to the selective pressures that arise in the context of competitive signal discrimination. That is, that these structures represent adaptations on the part of the maternal genome (encoding the maternal-store proteins) to the antagonistic coevolutionary environment that accompanies imprinting.

It may be possible to test this hypothesis as more is understood about the mechanisms of epigenetic regulation in early mammalian development. For example, if *AND*-linked methylation maintenance was driven by imprinting, we would not expect to find that Dnmt1o had evolved prior to the first imprinted genes. Data on the phylogenetic distribution of Dnmt1o is incomplete, and its trafficking behavior has been investigated only in mice. Similarly, we currently have only a partial understanding of how the active demethylation process varies among species. (For a discussion of the comparative aspects of these processes see, e.g. Morgan et al., 2005; Wilkins, 2005) Variation in these architectures among mammals may provide opportunities for hypothesis testing, particularly if these architectures covary with other features of the imprinting apparatus, or with relevant aspects of reproductive behavior.

In both of the contexts considered here—active demethylation and methylation maintenance—the evolutionary interests of the maternal-store proteins are expected to be aligned with those of the maternally derived loci. That is, regardless of whether selection on the *cis*-acting regulatory signals at imprinted loci is associative or evasive, theory would predict that it will be the paternally modified loci that will be under selection to mimic the maternally modified loci. This pattern of mimicry suggests other possible hypothesis tests. For example, once the imprinting status of many imprinted loci has been characterized across numerous mammalian species, loci could be identified where paternal epigenetic modifications have been lost in particular lineages. We might expect to find that the *cis*-acting elements associated with these loci bear a greater similarity to maternally modified loci in the lineages that have retained imprinting, compared with those lineages in which imprinting has been lost.

One challenge that awaits such a test is the fact that, in practice, it is not possible to categorize all imprinted loci simply as "paternally modified" or "maternally modified." Many imprinted loci have different modifications on each of the alleles. In some cases, there is a primary epigenetic mark that is established during gametogenesis, and numerous secondary modifications that occur later. The arguments presented here would apply strictly only to the primary mark and to its direct consequences. Other models will be required to analyse the selective forces acting on

these secondary modifications—particularly those that occur after transcriptional activation of the zygotic genome. In these cases, the maternal genome is responsible for the *trans*-acting factors that establish the primary mark, but the fetal genome (representing both the maternally derived and paternally derived alleles at an unimprinted locus) is responsible for the *trans*-acting factors that recognize the primary mark and establish a secondary mark in *cis*.

The discrimination problem in imprinting is further complicated by the presence of other classes of DNA sequences, such as retrotransposons, whose activity is regulated by an overlapping set of methylation mechanisms (Bestor and Bourc'his, 2004; Bourc'his and Bestor, 2004; Gaudet et al., 2004; Kaneda et al., 2004; Lane et al., 2003; Rakyan et al., 2003). The *cis*-acting elements associated with these loci will be subject to selective pressures that are not identical to those acting on any of the genetic factions discussed here (see, e.g. Bestor, 2003; Goll and Bestor, 2005; Wilkins, 2005). For example, under some circumstances, a retrotransposon might benefit from the loss of epigenetic marks that silence its transcription, even if all other loci involved (maternal, maternally derived, paternally derived, etc.) favored maintaining this silencing.

The overlap between the mechanisms of retrotransposon silencing and genomic imprinting creates another possible avenue for mimicry. Specifically, if the *cis*-acting elements associated with an epigenetically silenced allele were sufficiently similar to those associated with one or more retrotransposons, it might be impossible for the maternal-store regulatory machinery to reactivate that allele without simultaneously reactivating those retrotransposons. Because the maternal-store proteins will be under selection to reactivate paternally silenced, but not maternally silenced, alleles (Wilkins and Haig, 2002), the selective benefits of retrotransposon mimicry would apply specifically on paternally silenced loci. If this selective force has contributed significantly to the molecular evolution of imprinted gene regulation, then we might expect to find that paternally silenced loci bear a greater similarity to retrotransposable elements than maternally silenced loci do.

The arguments presented assume that the *trans*-acting factors present in the early embryo come entirely from the egg, and that they represent the mother's inclusive fitness interests. Recent evidence indicates that some enzymatic activities are contributed by the sperm, including phospholipase C (Swann et al., 2004), and 5-methylcytidyl deaminase (Jost et al., 2002). These results suggest that paternal *trans*-acting factors may have some capacity to influence epigenetic reprogramming after fertilization. Although it is likely that the impact of these factors is small compared with that of the maternal-store proteins, a complete understanding of the coevolutionary dynamics of this system might require consideration of yet another evolving genetic faction.

In the future, it will be possible to construct more realistic models that incorporate more of the complex interdependencies of fitness and molecular mechanism. In addition to the incorporation of elements such as retrotransposons and paternal *trans*-acting factors, future models might explicitly consider the evolution of novel regulatory architectures, as well as the molecular evolution of the DNA sequence motifs as they attempt to recruit or evade particular epigenetic modifications within this coevolutionary context. Constructing and interpreting such models in a meaningful way will depend on a more complete experimental characterization of the mechanisms of epigenetic gene regulation—and in particular how these mechanisms vary among loci and across species. However, the model considered here provides a framework for understanding certain recent observations in mammalian epigenetic reprogramming, and will hopefully be of use in designing and interpreting future experimental research.

Acknowledgements

Thanks to K. Eggan, D. Haig, D. Krakauer, C. Meiklejohn, and F. Úbeda for helpful discussions and to the anonymous reviewers for their comments on the manuscript.

Appendix A. Coevolutionary simulations

Each coevolutionary simulation included a population of 1000 receiver architectures and 10 populations of 1000 signals each. Each signal was represented as a string of 20 integers, each of which could take on values from zero through four. Each receiver system was likewise represented as a string of twenty integers valued zero through four. Each receiver architecture consisted of two receiver systems. Each generation in the simulation consisted of the following. Each architecture was compared against one signal from each of the 10 populations. These comparisons were chosen by sampling without replacement, so that each signal was compared against exactly one receiver architecture. Degree of recognition by a receiver system was determined by the number of matches between the receiver and signal strings of integers. This number of matches was determined for both of the receiver systems in each receiver architecture. The match between a signal and an architecture depends on the logical structure of the architecture. For an *AND*-linked architecture, the match is equal to the lesser of the two receiver system matches. For an *OR*-linked architecture, it is equal to the greater of the two.

Fitnesses were then assigned to each signal and architecture. The fitnesses assigned to the signals depended on the type of selection assumed. Under evasive selection, signals paid a fitness cost for matching the architecture, whereas under associative selection, they paid a fitness cost for failing to match the architecture. There were two classes of signals (S_1 and S_2 in the main text), and the 10 signal populations were divided into five populations of each type. Receiver architectures suffered a fitness cost for each signal of class S_1 that they failed to match, and a fitness

cost for each signal of class S_2 that they did match. Fitness costs to the architectures were multiplicative, with the maximum fitness for each interaction being one, and the minimum being 0.9. That is, an architecture that fully matched each S_1 signal (at all 20 sites), and completely failed to match each S_2 signal (matching at zero out of 20) would have a fitness of 1. An architecture that matched none of the S_1 signals, but matched all of the S_2 signals would have a fitness of $(0.9)^{10}$. An architecture that matched all 10 signals exactly, or completely failed to match any would have a fitness of $(0.9)^5$. Since each signal engages in only one interaction, the minimum fitness for a signal is 0.9.

For intermediate degrees of match between signal and receiver, the fitness value assigned to an interaction was determined from a hypergeometric function. Specifically, for a signal or receiver that benefits from matching, the fitness associated with matching at s sites was given by

$$0.9 + 0.1\left(\frac{\Gamma[b]}{\Gamma\left[b\,(m/L)\right]\Gamma\left[b\left(1-(m/L)\right)\right]}\frac{L}{bm}\left(\frac{s}{L}\right)^{b(m/L)} {}_2F_1\left[b\,\frac{m}{L}, 1+b\left(\frac{m}{L}-1\right), 1+b\,\frac{m}{L}, \frac{s}{L}\right]\right) \quad \text{(A1)}$$

and the fitness associated with a signal or receiver that suffers a fitness cost from matching is given by

$$1.0 - 0.1\left(\frac{\Gamma[b]}{\Gamma\left[b\,(m/L)\right]\Gamma\left[b\left(1-(m/L)\right)\right]}\frac{L}{bm}\left(\frac{s}{L}\right)^{b(m/L)} {}_2F_1\left[b\,\frac{m}{L}, 1+b\left(\frac{m}{L}-1\right), 1+b\,\frac{m}{L}, \frac{s}{L}\right]\right). \quad \text{(A2)}$$

The term m is the same as the m mentioned in the main text. Roughly speaking, it is the number of matches between the signal and receiver strings required to activate the receiver. More specifically, it is approximately equal to the value of s for which Eqs. (A1) and (A2) will equal 0.95. The term L is the length of the string, and is equal to 20 for all of the results presented here. The parameter b determines the steepness of the fitness function, with higher values of b corresponding to a steeper function. As b becomes very large, the function starts to look like a step function where the step occurs at m. For all results presented here b was set to 10. The hypergeometric function is represented as ${}_2F_1[a,b,c,d]$, following the notation used in *Mathematica*. Fig. A1 is a plot of Eq. (A1) as a function of s for two different values of m. The fitness values given by Eq. (A2) look similar, but with fitness being a decreasing, rather than an increasing, function of s.

The next generation for each population was then created by sampling with replacement, weighted by fitness. Specifically, an individual was chosen at random to reproduce. A random number between zero and one was then generated. If that number was smaller than the fitness of that individual, then the individual contributed an offspring to the next generation. This process was repeated until the entire next generation had been created. The new populations were then subjected to mutation. The number of mutations in each population was Poisson distributed with a mean of one. For each mutation, an individual and a site in the string of integers were chosen at random. The value at that site was then changed to a random value (0–4). For the signal populations, if the chosen site corresponded to a fixed site, no mutation occurred.

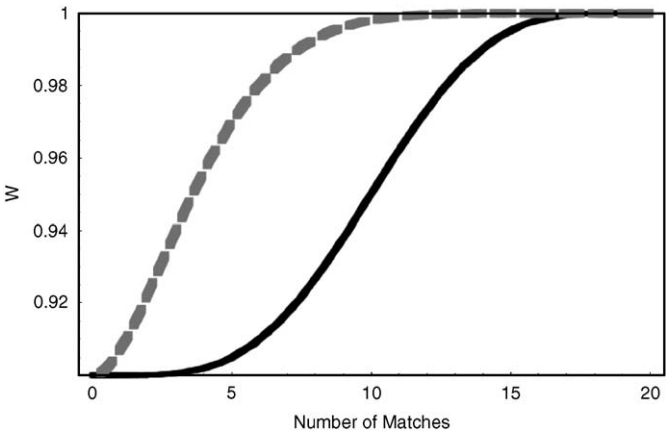

Fig. A1. The hypergeometric fitness function. The fitness function described by Eq. (A1) is plotted here as a function of s, the number of matches between signal and receiver. The fitness consequence, W, associated with a particular interaction (or failure to interact) is shown for two different match stringencies. The dashed gray line is the fitness function for $m = 4$, and the solid black line is for $m = 10$. In each case, the total string length, L, is 20, and the coefficient b, which determines the steepness of the transition between the low-fitness and high-fitness regions of the plot, was set to 10. These are the same values of L and b that were used in the simulations presented in the main text.

For simulations in which there are fixed sites on the signals ($f > 0$), f sites were first chosen at random for each of the 10 signal populations. Each of those sites was set to a random value (0–4). The same sets of fixed sites and values were used for the simulation runs with the *AND*-linked and *OR*-linked architectures. At the start of each simulation, each individual in each population was assigned a string of random values (0–4), with the exception of the fixed sites in the signal strings. Each simulation was run for 150,000 generations, and the first 100,000 were discarded, in order to assess the behavior of the architectures after the coevolutionary process equilibrated.

The simulations were implemented in a C program. The source code for the simulations is available from the author.

References

Bestor, T.H., 2003. Cytosine methylation mediates sexual conflict. Trends Genetics 19, 185–190.

Bestor, T.H., Bourc'his, D., 2004. Transposon silencing and imprint establishment in mammalian germ cells. Cold Spring Harbor Symp. Quant. Biol. 69, 381–387.

Bhattacharya, S.K., Ramchandani, S., Cervoni, N., Szyf, M., 1999. A mammalian protein with specific demethylase activity for mCpG DNA. Nature 397, 579–583.

Boeke, J., Ammerpohl, O., Kegel, S., Moehren, U., Renkawitz, R., 2000. The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A. J. Biol. Chem. 275, 34963–34967.

Bourc'his, D., Bestor, T.H., 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 431, 96–99.

Burt, A., Trivers, R., 1998. Genetic conflicts in genomic imprinting. Proc. R. Soc. London Ser. B-Biol. Sci. 265, 2393–2397.

Cedar, H., Verdine, G.L., 1999. Gene expression—the amazing demethylase. Nature 397, 568–569.

Dean, W., Santos, F., Stojkovic, M., Zakhartchenko, V., Walter, J., et al., 2001. Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. Proc. Natl. Acad. Sci. USA 98, 13734–13738.

Delaval, K., Feil, R., 2004. Epigenetic regulation of mammalian genomic imprinting. Curr. Opin. Genet. Dev. 14, 188–195.

Doherty, A.S., Bartolomei, M.S., Schultz, R.M., 2002. Regulation of stage-specific nuclear translocation of Dnmt1o during preimplantation mouse development. Dev. Biol. 242, 255–266.

Gaudet, F., Rideout, W.M., Meissner, A., Dausman, J., Leonhardt, H., et al., 2004. Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. Mol. Cell. Biol. 24, 1640–1648.

Goll, M.G., Bestor, T.H., 2005. Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. 74, 481–514.

Haig, D., 1993. Genetic conflicts in human-pregnancy. Quart. Rev. Biol. 68, 495–532.

Haig, D., 1996. Placental hormones, genomic imprinting, and maternal-fetal communication. J. Evol. Biol. 9, 357–380.

Haig, D., 1997. Parental antagonism, relatedness asymmetries, and genomic imprinting. Proc. R. Soc. London Ser. B-Biol. Sci. 264, 1657–1662.

Haig, D., 2000. The kinship theory of genomic imprinting. Annu. Rev. Ecol. System. 31, 9–32.

Haig, D., 2004. Genomic imprinting and kinship: how good is the evidence? Ann. Rev. Genet. 38, 553–585.

Haig, D., Westoby, M., 1989. Parent-specific gene-expression and the triploid endosperm. Am. Nat. 134, 147–155.

Howell, C.Y., Bestor, T.H., Ding, F., Latham, K.E., Mertineit, C., et al., 2001. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell 104, 829–838.

Jaenisch, R., Bird, A., 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. 33, 245–254.

Jost, J.P., Thiry, S., Siegmann, M., 2002. 5-methyldeoxycytidine monophosphate deaminase and 5-methylcytidyl-DNA deaminase activities are present in human mature sperm cells. FEBS Lett. 519, 128–134.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., et al., 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429, 900–903.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., et al., 2003. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis 35, 88–93.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., Haaf, T., 2000. Embryogenesis—demethylation of the zygotic paternal genome. Nature 403, 501–502.

Mochizuki, A., Takeda, Y., Iwasa, Y., 1996. The evolution of genomic imprinting. Genetics 144, 1283–1295.

Moore, T., Reik, W., 1996. Genetic conflict in early development: parental imprinting in normal and abnormal growth. Rev. Reproduct. 1, 73–77.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W., Petersen-Mahrt, S.K., 2004. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues—implications for epigenetic reprogramming. J. Biol. Chem. 279, 52353–52360.

Morgan, H.D., Santos, F., Green, K., Dean, W., Reik, W., 2005. Epigenetic reprogramming in mammals. Hum. Mol. Genet. 14, R47–R58.

Morison, I.M., Ramsay, J.P., Spencer, H.G., 2005. A census of mammalian imprinting. Trends Genet. 21, 457–465.

Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., et al., 1999. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat. Genet. 23, 58–61.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., et al., 2000. Active demethylation of the paternal genome in the mouse zygote. Curr. Biol. 10, 475–478.

Rakyan, V.K., Chong, S., Champ, M.E., Cuthbert, P.C., Morgan, H.D., et al., 2003. Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. Proc. Natl. Acad. Sci. USA 100, 2538–2543.

Ratnam, S., Mertineit, C., Ding, F., Howell, C.Y., Clarke, H.J., et al., 2002. Dynamics of Dnmt1 methyltransferase expression and intracellular localization during oogenesis and preimplantation development. Dev. Biol. 245, 304–314.

Reik, W., Walter, J., 2001. Evolution of imprinting mechanisms: the battle of the sexes begins in the zygote. Nat. Genet. 27, 255–256.

Santos, F., Hendrich, B., Reik, W., Dean, W., 2002. Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev. Biol. 241, 172–182.

Spencer, H.G., Williams, J.M., 1997. The evolution of genomic imprinting: Two modifier-locus models. Theor. Popul. Biol. 51, 23–35.

Swann, K., Larman, M.G., Saunders, C.M., Lai, F.A., 2004. The cytosolic sperm factor that triggers $Ca2^{+}$ oscillations and egg activation in mammals is a novel phospholipase C: PLC zeta. Reproduction 127, 431–439.

Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T., Sedgwick, B., 2002. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. Nature 419, 174–178.

Vairapandi, M., 2004. Characterization of DNA demethylation in normal and cancerous cell lines and the regulatory role of cell cycle proteins in human DNA demethylase activity. J. Cell. Biochem. 91, 572–583.

Vairapandi, M., Liebermann, D.A., Hoffman, B., Duker, N.J., 2000. Human DNA-demethylating activity: a glycosylase associated with RNA and PCNA. J. Cell. Biochem. 79, 249–260.

Wade, P.A., Gegonne, A., Jones, P.L., Ballestar, E., Aubry, F., et al., 1999. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat. Genet. 23, 62–66.

Wilkins, J.F., 2005. Genomic imprinting and methylation: epigenetic canalization and conflict. Trends Genet. 21, 356–365.

Wilkins, J.F., Haig, D., 2001. Genomic imprinting of two antagonistic loci. Proc. R. Soc. London Ser. B-Biol. Sci. 268, 1861–1867.

Wilkins, J.F., Haig, D., 2002. Parental modifiers, antisense transcripts and loss of imprinting. Proc. R. Soc. London Ser. B-Biol. Sci. 269, 1841–1846.

Wilkins, J.F., Haig, D., 2003. What good is genomic imprinting: the function of parent-specific gene expression. Nat. Rev. Genet. 4, 359–368.