EVOLUTIONARY EPIGENETIC THEORY

by

Root Gorelick

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

ARIZONA STATE UNIVERSITY

May 2004

EVOLUTIONARY EPIGENETIC TEHORY

by

Root Gorelick

has been approved

May 2004

APPROVED:

_____, Chair

_____

_____

_____

_____

Supervisory Committee

ACCEPTED:

_____

Director of the School

_____

Dean, Graduate College

ABSTRACT

Epigenetic effects are important in evolutionary biology. Yet, except in evolutionary developmental biology, there is little or no evolutionary epigenetic theory. Part of the problem has been not realizing how readily evolutionary epigenetics fits into existing formulations of quantitative and population genetics nor when it does not. Part of the problem has been delineating definitions of epigenetic signals and effects, i.e. reconciling developmental definitions with molecular definitions regarding cytosine methylation and chromatin formation. From a molecular perspective, epigenetic signals are heritable signals other than DNA nucleotides. From an ontogenetic perspective, epigenetic signals are those that affect development. To bridge this gap, I first develop several quantitative and population genetic models that include epigenetic effects and then develop applications that arise from these models. These model the relative fluidity of epigenetic signals and effects, and epigenetic changes induced by gene duplications. Altogether this dissertation presents nine key findings and conceptual advances. (1) Formal Fisherian models confirm that epigenetic effects can be considered equivalent to epistatic effects. (2) Qualitatively, however, epigenetic signals are more fluid than DNA nucleotides, a distinction that can be integrated into quantitative and population genetic models. (3) Regression models show explicitly how phenotypic plasticity can arise from fluidity of epigenetic signals. (4) Environmental shocks alter epigenetic signatures, thereby affecting development and disease. (5) Origins of dioecy and sex chromosomes may be evolutionary artifacts of cytosine methylation. (6) Derived lineages have had more gene duplications and hence have shorter sex chromosomes. (7) Reset of epigenetic signatures

each generation may be the *sine qua non* of meiosis. (8) Polyploidy may epigenetically trigger heterochrony and thereby also macroevolutionary radiations. (9) Linkage disequilibrium is a genetic decomposition that is equivalent to integer decomposition. In conclusion, there is enormous potential for advancing evolutionary theory by meshing epigenetics with population and quantitative genetics.

"Neo–Darwinism involves a breach between organism and nature as complete as the Cartesian dualism of mind and matter; an epigenetic consideration of evolution goes some way towards healing it." (Waddington, 1957: ix)

"Without question, the genetic apparatus is the guarantor of the basic stability of genetic information. But equally without question, it is a more complex system, with more complex forms of feedback, than had been previously thought. Perhaps the future will show that its internal complexity is such as to enable it not only to program the life cycle of the organism, with fidelity to past and future generations, but also to reprogram itself when exposed to sufficient environmental stress – thereby effecting a kind of 'learning' from the organism's experience. Such a picture would be radical indeed, and it would be one that would do justice to McClintock's vision: it would imply a concept of genetic variation that is neither random nor purposive – and an understanding of evolution transcending that of both Lamarck and Darwin." (Keller, 1983: 194-195)

"It has been concluded that the changed phenotypic expressions of such loci are related to changes in a chromatin element other than that composing the genes themselves, and that mutable loci arise when such chromatin is inserted adjacent to the genes that are showing variegated expression." (McClintock, 1950: 347)

"[The] most obviously fruitful role [of theory] is in providing explicit direction for research. From theory we can deduce conclusions not previously reached and that are occasionally counterintuitive…. If fortune smiles on the theorist, the theoretically derived conclusion will be readily testable by observation…. Theory can also be useful, if less directly, if it merely clarifies relations among concepts, even without explicit reference to testable hypotheses. It helps put our conceptual house in order and prevent self-deception." (Williams, 1988: 297)

"A work on [evolutionary theory], placed into our hands, is apt to be experienced as a bullet – either to be dodged, or loaded and fired at unbelievers. Skim the conclusions and decide which. One is, of course, thereby safe and sound when confronted by an enemy's projectile; and pluralism produces a surfeit of enemies. But the price of safety is impotence. [Evolutionary] thought is most likely to influence others when it forces its *proponent* to accept conclusions found personally distasteful. By limiting my autonomy, binding myself to conclusions I dislike, I am less dangerous to others – and, perhaps, more likely to find common ground with these others elsewhere." (Pollock, 2001: 8)

"Writing is hard work. It is boring and lonely. And there are too many long stretches of panic and self-hatred between the moments of inspiration. I have never been able to endure this drudgery and finish a piece that I did not care about passionately." (Califia, 1988: 10)

ACKNOWLEDGEMENTS

First and foremost, thanks to Sue Bertram and Manfred Laubichler. I have pestered them incessantly about the ideas herein and about many other technical and non-technical matters. They have patiently listened to my incoherent rants and were encouraging and helpful throughout. Had it not been for Sue and Manfred, this dissertation would have been a pale shadow of itself.

Thanks to my committee members – Manfred Laubichler (chair), Jennifer Fewell, Tom Dowling, Tom Taylor, John Sabo – for your patience and help. Also thanks to Greg Pollock who was officially removed from my committee for spurious administrative reasons. Thanks to my previous graduate school advisors – V.S. Krishnan (mathematics), Dan Howard (biology), Doug Gegax (economics), Ed Schlee (economics), Tom Dowling (biology), and Jennifer Fewell (biology) – who probably always believed I would finish a PhD, although none of them were probably ever sure when nor in what subject area.

Thanks to the graduate college and biology department for giving me summer research funding for all three years that I was studying biology at Arizona State University.

Thanks to the Arizona State University Library staff and holdings. My work relied on having access to a wealth of literature.

Three people helped convince me to look at the world through epigenetic-colored glasses. Shirley Tilghman introduced me to cytosine methylation, although she would barely recognize where I have run with her ideas. Once I began, Greg Pollock and Roy Osborne kept prodding me along, always asking me for the essential nagging details that

any decent theory rests on.

Thanks to Marakatha and V.S. Krishnan who have seen me through career, philosophical, and personal changes for over two decades now. They have always opened up their home and hearts to me as though I were their son.

Finally, a special thanks to Hunter Raven Gorelick. I will probably never know whether this was coincidental or not, but my productivity and creativity increased by an order of magnitude the moment you were born and amazingly has continued to increase ever since. You and Sue kept me energized for the past four years and hopefully for many more.

TABLE OF CONTENTS

Page

Page

CHAPTER 1

INTRODUCTION

In this introduction, I first describe what is meant by epigenetics. I define *epigenetic* in two ways: (1) heritable signals other than those coded for by DNA nucleotides and (2) fluid genetic signals, i.e. those with lower heritabilities than DNA nucleotides (if one considers nucleotides to be the phenotype). These above two definitions of epigenetic effects establish the equivalence between epigenetic and epistatic effects within quantitative genetics and allow for a generalization of indirect genetic effects that models environmentally-alterable meitocally-heritable effects, respectively. Next, I discuss several practical consequences of this evolutionary epigenetic theory. These consequences are diverse, ranging from inducing sex change in plants to preventing cancer in humans. Finally, I provide a roadmap to the chapters, which are clustered into two (overlapping) parts. Part I of this dissertation contains quantitative and population models for analyzing the evolutionary consequences of epigenetic phenomena. Part II contains evolutionary consequences of those models and epigenetic changes induced by gene duplications. Part I is largely mathematical; Part II is not.

The term epigenetic was originally used to describe all factors controlling gene expression and cell differentiation. Waddington (1940) used the term epigenetic as an amalgamation of epigenesis and genetics, where epigenesis refers to the gradual and progressive development of new structures (Hall, 1992). It was not until after Waddington was writing his seminal book that biologists realized that DNA is primarily

contained in chromosomes (Caspersson & Schultz, 1940 citing Brachet, 1940) and that the information coding portion of genes is DNA (Avery, MacLeod & McCarty, 1944; Hershey & Chase, 1952). Genetic phenomena, like epigenetic ones, were originally considered to be of unknown molecular cause. Only genetic and epigenetic *effects* were known. Fisher defined what we now call additive genetic variance as the proportion of phenotypic variance that gets transmitted from one generation to the next (Fisher, 1930). Note that this is a population level definition of genetic effects because it requires one to examine average effects and variances. In contrast, Waddington used the term epigenetic to refer to any phenomena that affected development *within an individual* and, therefore implicitly, *within a generation* (although, he also sometimes used the term epigenetic to refer to trans-generational phenomena; Waddington, 1957).

The meaning of the term genetic changed with the introduction of the central dogma and sequence hypothesis (Crick, 1970), what could be termed a preformist view of molecular biology (Hall, 1992). This provided a one-way path from DNA nucleotide sequences, to messenger RNA (transcription), to protein synthesis (translation), where the nucleotides in eggs and sperm form a molecular homunculus. Because of this paradigm, molecular biologists effectively re-defined the term genetic to mean DNA nucleotide sequences; a definition that many quantitative geneticists would admittedly find unpalatable.

Use of the term epigenetics also changed in the 1970s as molecular biologists began proposing mechanisms, especially cytosine methylation, by which development and gene regulation could be controlled (Holliday & Pugh, 1975; Riggs, 1975). Many of

these molecular signatures gradually change irreversibly as cells and tissues differentiate and age. Therefore, whichever molecular mechanisms control development must be reset each generation. "This suggests a supplementary definition of epigenetics to include transmission from one generation to the next, other than the DNA sequence itself" (Holliday, 1994: 454). Holliday thereby switched the focus of epigenetics from strictly intra-generational to inter-generational (also see Holliday, 1987). Therefore, if there is any ambiguity, I use the terms 'mitotically-heritable' and 'meiotically-heritable' to distinguish the two different contemporary meanings of the term epigenetic. The awkward adjectival phrase 'meiotically-heritable' can be a necessity because much of the medical literature, especially in cancer research, uses the term heritable to mean signals that are transmitted through mitosis, without ever mentioning or caring whether such signals get transmitted through meiosis and syngamy. Transmission through syngamy is non-trivial. Cytosine methylation and other epigenetic signals are usually reset in both meiosis and syngamy (although not always completely). The classic example is genomic imprinting, where female and male homologues have different methylation or chromatin patterns. During syngamy, many female genomic imprints often (but not always) over-write the homologous male imprint (Mayer, Niveleau, Walter, Fundele & Haaf, 2000; Mayer, Smith, Fundele & Haaf, 2000; Oswald, Engemann, Lane, Mayer, Olek, Fundele, Dean, Reik & Walter, 2000). Meiotically-heritable epigenetic signals include cytosine methylation, chromatin structure, and histone acetylation and undoubtedly also include other molecular signals, such as those mediating RNA editing. Mitotically-heritable epigenetic signals include all the same molecular mechanisms that are meiotically

inherited, but also possibly others that cannot be transmitted through meiosis and syngamy. As can be seen by comparing asexual taxa with sister taxa that are sexual but lack outcrossing, meiosis and syngamy provide a partial roadblock to transmission of some traits. When there is no confusion, I will simply use the term heritable to mean meiotically-heritable.

Holliday (1994) defined epigenetic as any signals that are heritable through meiosis and syngamy except those directly encoded by DNA nucleotides. From a quantitative genetic perspective, therefore, epigenetic and genetic are synonymous. However, epigenetic signals sensu Holliday are not as heritable as nucleotide sequences. It is much easier to alter epigenetic signals such as cytosine methylation and chromatin formation with environmental perturbations than it is to alter nucleotide sequences. Therefore, I define epigenetic signals as fluid genetic signals. The degree of this epigenetic fluidity – the flip side of which is lower heritability – has not been quantified for any epigenetic signals. That said, there exists a consensus that offspring epigenetic signals qualitatively resemble that of their parents and that the faithfulness of this transmission is less than that of nucleotides. Surprisingly, though, nobody has explicitly measured the heritability of cytosine methylation, chromatin formation, or any form of genomic imprinting. All that has been noted is that qualitatively these signals seem to be transmitted from one generation to the next. Even less is known about heritability of other epigenetic signals and effects, such as ribosomal RNA (rRNA), microtubule organizing centers (MTOCs), transposons, developmental modules, and standing waves of enzymes. There is not too much difficulty in picturing how rRNA could be heritable,

as it is thought to be the precursor of DNA in biological systems (Cech & Bass, 1986; Gilbert, 1986). In fact, cytosine methylation is probably nothing more than an artifact of this RNA world – thymine is nothing more than methylated uracil (Poole, Penny & Sjöberg, 2000; Poole, Penny & Sjöberg, 2001). MTOCs appear to be inherited along with other sub-cellular machinery (maybe including standing waves of enzymes?) in egg cells.

There has been a constant epigenetic thread running through evolutionary biology since the discovery of position effect variegation in the 1920s (Sturtevant, 1925), transposons in the 1940s (McClintock, 1950), genomic imprinting in the 1960s (Crouse, 1960), and the regulatory roles of cytosine methylation in the 1970s (Holliday & Pugh, 1975). My guess is that there are three reasons why these epigenetic phenomena have not been incorporated into evolutionary theory. First, prior to the 1940 publication of Waddington's classic book on epigenetic effects and Caspersson and Schultz's paper showing that chromosomes contain DNA, many epigenetic phenomena may have been considered equivalent to genetic ones (i.e. Fisher's black box approach that subsumed things that today we would consider genetic and epigenetic), I resuscitate this line of thinking in chapter 2. Second, people working on epigenetic signals seemed to be more interested in their molecular and genetic aspects than in evolutionary theory *per se*. Third, the central dogma and sequence hypothesis (Crick, 1970) effectively allocate epigenetic phenomena to a minor evolutionary role. Thus far, quantitative genetics has only been used to estimate developmental epigenetics (Cowley & Atchley, 1992; Wolf, Frankino, Agrawal, Brodie & Moore, 2001), without mention of the other

aforementioned meiotically-heritable epigenetic phenomena. Cowley and Atchley (1992) and Wolf et al. (2001) each took the modular approach typical of modern evolutionary developmental biology, where the arrows in their path analyses go from one module or developmental stage to another, rather than the arrows going from one generation to another.

Study of epigenetic effects has largely been the bailiwick of developmental biologists. This dissertation touches on those developmental aspects, especially in Appendices B and C, where I explicitly discuss the roles of cytosine methylation in cell differentiation and de-differentiation in animals and plants, respectively. There are, however, much deeper connections between molecular forms of epigenetic signals (including cytosine methylation) and more traditional views of developmental biology. In chapter 2, I show how epigenetic effects are mathematically equivalent to physiological epistatic effects. In the discussion, I also describe outlines of my theory that suggest how polyploidy (and possibly less extensive gene duplications) can be the proximate cause for neoteny and paedomorphism.

Failure to recognize epigenetic inheritance can lead to erroneous or unfounded inferences. When examining only phenotypic data, traits coded by alleles on sex chromosomes cannot be distinguished from those caused by genomic imprinting on autosomes. A typical example of this failure to recognize epigenetic effects is found in Iyengar, et al. (2002), who showed heritable phenotypic traits that were paternally inherited in a moth. From this, they concluded that these phenotypic traits are encoded by an allele on the Z chromosome (the Z chromosome in ZW lineages corresponds with

the X chromosome in XY lineages). However, they failed to mention that all moths and butterflies (Lepidoptera) contain genomic imprinting. In fact, genomic imprints are what distinguish lepidopteran Z from W chromosomes (Traut & Marec, 1996). Putative genomic imprinting causing the phenotypic differences seen in these moths could be on autosomes. Because of the paradigm that all heritable traits have a DNA nucleotide sequence basis, Iyengar *et al*. (2002) seemed to have been unaware of the equally viable alternative hypothesis that these phenotypic traits may have an epigenetic basis.

A large reason for epigenetic signals being ignored are the divergent definitions for the terms 'gene' and 'genetic' by those studying molecular biology, quantitative genetics, and developmental biology (Kitchner, 1992; Sarkar, 1998; Portin, 2002). Epigenetic signals such as cytosine methylation or chromatin formation that regulate gene expression would be considered 'genetic' by the *cis-trans* test, essentially being considered genetic enhancers, even though they may be non-Mendelian and may be heritably altered by the environment (Benzer, 1959; Serfling, Jasin & Schaffner, 1985). Molecular biologists consider epigenetic signals to be any "heritable changes in gene function that cannot be explained by changes in DNA sequence" (Riggs, Martienssen & Russo, 1996b: 1). Quantitative geneticists, however, consider genes and genetic signals to be any portion of variation that cannot be explained by environmental variation. Quantitative geneticists would therefore consider genomic imprinting to be a non-Mendelian genetic effect and environmentally-alterable meiotically-heritable epigenetic signals to be a form of genotype-by-environment interaction.

Throughout this dissertation, I take the quantitative genetic perspective that 'genetic' is equivalent to 'heritable' (Lush, 1937), a definition that several classical and molecular geneticists would admittedly find unpalatable (e.g. Sarkar, 1998). To be precise, I confine attention to narrow sense heritability, so that we are only looking at additive effects and do not confound matters with dominance and interaction effects (that are aggregated in with additive effects to compute broad sense heritability). Although this definition has well-documented problems, it has also provided the foundations upon which much evolutionary theory has been developed over the past seventy years. In particular, equating genetic with heritable provides a black box approach to evolutionary modelling, which is extraordinarily useful when the underlying molecular genetic and epigenetic mechanisms are unknown or unspecified. Thus, evolutionary epigenetic can utilize existing evolutionary genetic theory.

In chapter 2, I take the position that epigenetic effects are simply modifier loci. They can thus be incorporated into a Fisherian model of epistasis. Apparently this idea has never been explicitly proffered in quantitative genetics, although it seems implicit even in the work of Fisher (1930). By modelling epigenetic effects as a form of physiological epistasis, I have placed epigenetics into the Fisherian/neo-Darwinian formulation of evolutionary theory. Chapter 2 therefore contains a multi-locus model of physiological epistasis.

In addition to the more obvious developmental epigenetic signals, sex-limited traits also fit nicely into the multi-locus models of epistasis/epigenetics, especially if we invoke the theoretical frameworks of Lande (1980) or Bulmer (1972). Such sex-limited

epigenetic effects include those examples of genomic imprinting that seem highly conserved from one generation to the next (i.e. are highly heritable).

From both a theoretical and an empirical perspective, however, a physiological model of epistasis cannot subsume all forms of epigenetic effects for several reasons. Any multi-locus model of physiological epistasis will contain too many terms and coefficients (i.e. too many interactions between various loci) to be useful to empirical quantitative geneticists performing heritability experiments. In part, that is why Fisher amalgamated the effects of many discrete loci into a single term that aggregates a continuum of loci each with infinitesimal effect. But an even more fundamental problem arises in that physiological epistasis only models direct genetic effects. In chapter 4 and Appendix A, I argue that the lower heritability of epigenetic effects – at least compared with nucleotide sequences – is largely due to epigenetic signals being heritably altered by environmental perturbations. The resulting epigenetic fluidity at the hands of environmental perturbations can be modelled as a generalization of indirect genetic effects (Roff, 1986). These fit uncomfortably into the maternal effects framework of Kirkpatrick and Lande (1989), but more are comfortably placed into the macro-environmental condition framework of Bulmer (1972). Therefore, in chapter 4, I show how to model these fluid epigenetic effects, but only with continuous traits, and not with discrete multi-locus traits.

From an evolutionary perspective, chapters 2 and 4 are extraordinarily important insofar as they provide new routes by which epigenetic effects contribute to additive genetic variation. One of the quandaries in evolutionary biology has been to explain the

existence of substantial additive genetic variation in light of Fisher's fundamental theorem of natural selection (Fisher, 1930). Chapter 2 partly obviates this problem by showing that epigenetic effects are equivalent to epistatic effects. Epistasis is known to provide additive genetic variance and even convert dominance effects at a locus into additive effects over the entire genome (Goodnight, 2000). Chapter 4 partly obviates this problem of lack of additive genetic variance by generalizing the concept of indirect genetic effects to the additive effects of environmentally-alterable epigenetic effects (what Waddington, 1957 called 'environmentally-inducible gene mutations'). These two new sources of additive genetic variance arise from constitutive and facultative epigenetic effects, i.e. the non-fluid and fluid aspects of epigenetic effects, respectively.

In chapter 5 and the appendices, I discuss evolutionary epigenetic theory that arises from the above models and two features of many epigenetic signals that qualitatively set them apart from genetic ones: relative epigenetic fluidity (discussed in the next paragraph) and the triggering of meiotically-heritable epigenetic changes due to gene duplications (discussed in the next page).

Epigenetic signals, especially those attached to DNA nucleotides (e.g. cytosine methylation and chromatin formation), are more fluid than their underlying nucleotides. This epigenetic fluidity occurs on two separate time scales: the life of an individual organism (ontogeny) and over evolutionary time (roughly phylogeny). For an example of epigenetic fluidity over ontogeny, chromatin bands vary in a predictable fashion that reflects differing gene activity during developmental stages of ontogeny (Ficq & Pavan, 1957). Nonetheless, these epigenetic signals are highly heritable, albeit not as heritable

as nucleotide sequences themselves. In other words, even on an evolutionary time scale, epigenetic signals are more fluid than nucleotide signals. As I discuss in the next paragraph, epigenetic fluidity over evolutionary time is partly due to epigenetic changes caused by gene duplications. Relative epigenetic fluidity provides a counterpoint between signals that molecular biologists would call epigenetic and those signals such as nucleotides that they would call genetic. Thus, even though I show that from a purely Fisherian perspective, genetic and epigenetic are equivalent, epigenetic fluidity, at both intra- and inter-generational time scales, sets evolutionary epigenetic theory qualitatively apart from evolutionary genetic theory.

Gene duplications are known to be an important evolutionary force. Gene duplications have played roles in speciation insofar as polyploidy and duplication of homeotic genes are believed to be major causes of speciation and morphological differentiation. Gene duplications have proven extraordinarily valuable in inferring phylogenies (Griffith, 2004). Gene duplications also include transposons, which are often thought of as genomic parasites that co-evolve with their hosts. All of these types of gene duplications are known to trigger epigenetic changes. However, nobody has yet explored very deeply their evolutionary epigenetic consequences. Providing the evolutionary theory and consequences of those epigenetic changes is an important new suggestion in this dissertation.

As an aside, there is very much a chicken-and-egg problem with gene duplications and epigenetic changes. They both seem to trigger each other. There is plenty of evidence that gene duplications trigger addition of cytosine methylation and

heterochromatin formation; see Appendix B, C, and D for citations. There is also circumstantial evidence that loss of cytosine methylation or heterochromatin at a classical genetic locus makes that locus more susceptible to gene duplications, inversions, or insertion of transposons (Waugh O'Neill, O'Neill & Graves, 1998; Tuck-Muller, Narayan, Tsien, Smeets, Sawyer, Fiala, Sohn & Ehrlich, 2000; Waugh O'Neill, Eldridge & Graves, 2001). Thus, gene duplications could be considered epigenetic phenomena. However, for this dissertation, I simply consider gene duplications to be an exogenous environmental factor that triggers epigenetic change.

Throughout this dissertation I confine attention to eukaryotes, with particular emphasis on those lineages that undergo meiosis (possibly automixis). The mathematical models developed in chapters 2, 3, 4, and Appendix B assume that gametes are haploid and somatic cells are diploid, although this could be extended to higher ploidy levels. The applications contained in chapter 5 and the appendices specifically deal with a wide swath of organisms from sexual fungi, to flowering and non-flowering plants, to marine and terrestrial invertebrates, to vertebrates, including humans.

DETAILED ROADMAP OF DISSERTATION CHAPTERS

By realizing that epigenetic effects are equivalent to epistatic effects, I incorporate epigenetic effects into a Fisherian/neo-Darwinian framework by ‘simply’ building a multi-locus model of epistasis. Until now, models of physiological epistasis only existed for pairs of bi-allelic loci.

Fisher devised quantitative genetics before anybody knew that genes were comprised of DNA. Thus his methods had to accommodate all forms of heritable signals, from DNA nucleotides, to ribosomal RNA, to cytosine methylation, to chromatin formation. That is, Fisher's methods should be equally valid for signals that today we consider either genetic or epigenetic. I conceptualize an epilocus as simply another locus. For epigenetic signals such as cytosine methylation, the presence or absence of methylation at a nucleotide locus can be thought of as simply another very tightly linked locus, albeit effectively a regulatory locus. Likewise, presence or absence of heterochromatin or histone acetylation on a classical genetic locus (roughly a stretch of DNA of a thousand or so nucleotides) can be thought of as an epilocus that is tightly linked to the classic genetic locus. Therefore, to build a formal Fisherian model of epigenetic effects that includes ordinary physiological epistasis between say a pair of genetic loci, we need to build a four-locus model of physiological epistasis. In chapter 2, I build this model for an arbitrary number of loci.

One complication arose in building the multi-locus model of physiological epistasis. In order to define average effects of an allele, we first need to define linkage disequilibrium for an arbitrary number of loci. Sixty years ago, Geiringer sketched how to do this, but never gave an explicit formula. In chapter 3, I finish up the work she began and, in so doing, also link her definition of linkage disequilibrium with another definition based upon probabilistic independence. The keys to my work here are (1) properly defining linkage disequilibrium for a single locus and (2) realizing that linkage disequilibrium is a genetic decomposition that happens to be equivalent to the number-

theoretic decomposition of a positive integer into its additive parts. Immediately following this technical diversion of chapter 3, I delve back into epigenetic theory.

Chapter 4 provides a bridge between pure theory and applications. The formalism established in chapters 2 and 3 can be used by theorists, but is not the sort of tool that can readily be pulled off the shelf by empirical evolutionary biologists. Therefore, in chapter 4, I provide a simple quantitative genetic model for the statistical estimation of heritability of epigenetic signals, i.e. decomposing phenotypic variance into components including those due to epigenetic effects. Epigenetic effects provide a possible mechanism underlying genotype-by-environment (GxE) interactions and phenotypic plasticity, phenomena that have hitherto largely been conceptualized as residual statistical non-linearities. This model places environmentally-alterable meiotically-heritable epigenetic effects squarely within the Fisherian/neo-Darwinian framework.

The tools developed in chapter 4 are more than mere convenience, however. Although epigenetic effects can formally be modeled as genetic effects, there is something qualitatively different about them. DNA nucleotides are rather permanent edifices. They do not incur very many mutations of any form (point mutations, insertions, deletions, etc.) over evolutionary time or over ontogeny of an individual. By contrast, Waddington used the term epigenetic to refer to any phenomena that affected development within an individual and, therefore implicitly, within a generation. Molecular biologists also think of many epigenetic signals changing over the life of an individual, hence their distinction between facultative and constitutive heterochromatin.

Epigenetic signals also appear to be more fluid than DNA nucleotides over evolutionary time.

Consider cytosine methylation and chromatin formation, which reside on top of DNA nucleotides ("epi-" is Greek for "on top of"). This is analogous to an electrical wire, with DNA nucleotides forming the copper conducting portion, the epigenetic signals forming a heterogeneous layer of insulation, and the entire genome forming a bundled cable. The insulation can only act by altering the electrical and magnetic signals flowing through the wire (gene regulation), especially if the insulation is thin in places and there is shorting and cross talk between individual wires in the bundled cable Ironically this is the very role that Jack Schultz had in 1937 had originally ascribed to DNA: "[C]oncentrations of DNA along the chromosome acted to block the expression of expression of the nearest gene" (Judson, 1979: 235). Furthermore, it is much easier to damage the insulation (epigenetic signatures), say with a knife or heat, than it is to damage the underlying copper conducting wire (DNA nucleotides). Some of the environmental damage to epigenetic signals is corrected during meiosis and syngamy, but many of the epigenetic changes are transmitted to the next generation. Another way that this analogy works is that as bundled cables age, the insulation around the wires changes/degrades in a fairly predicable way, whereas the copper conducting portion of the wire seems relatively unaffected.

Cytosine methylation and chromatin formation can also be thought of as redaction of text in a four-letter alphabet. The more methylation or chromatin, the less readily the underlying DNA sequence can be read (transcribed). This is essentially the manner in

which these epigenetic signals regulate development, as well as disease susceptibility (Appendix A), sex determination and dioecy (Appendix B and C).

Regardless of analogies, epigenetic signals change much more readily over both ontogeny and phylogeny than do nucleotides. One of my goals in chapter 4 is to account for this epigenetic fluidity over evolutionary time (epigenetic fluidity over ontogeny has previously been modeled by (Cowley & Atchley, 1992 and Wolf et al., 2001). In chapter 4, epigenetic effects only enter into decompositions of additive genetic variance when they are modified by environmental perturbations. The contributions of constitutive epigenetic signals (e.g. much of genomic imprinting) to additive genetic variance have to be computed using a separate methodology, which I am developing separately (Gorelick & Bertram, submitted).

This dissertation is not just about theoretical models, but is also meant to cover a panoramic range of evolutionary epigenetic theory. Thus, the appendices contain evolutionary implications of the relative epigenetic fluidity and changes in epigenetic signals that are triggered by gene duplications. Appendix A provides explicit theoretical applications in human medicine and environmental policy for the mathematics derived in chapter 4. Appendix B, C, and D deal with evolutionary epigenetic implications of gene duplications for the origins and maintenance of sex. Appendix A, Appendix C, and chapter 5 also delve explicitly into the roles of fluid epigenetic signals in developmental biology. It would seem peculiar indeed to write about epigenetic effects without mention of development.

Appendix A directly applies the model of chapter 4 to human medicine. I propose that environmental perturbations can cause substantial heritable epigenetic changes to a lineage. I provide evidence from the literature regarding environmentally-induced heritable epigenetic changes. I then show how this could change our paradigms about medicine, debunking the nature versus nurture dichotomy (cf. Hogben, 1933) that is so pervasive in human medicine and changing emphasis from prevention to cure of human diseases. Once contaminants have adversely altered an individual's epigenetic signals, this harm will be transmitted to future generations even if they are not exposed to the contaminant. Exposure to environmental shocks such as free radicals or other carcinogens can alter cytosine methylation patterns on regulatory genes. This can cause cancer by up-regulating genes for cell division or by down-regulating tumor suppressor genes. I also highlight how such a paradigm shift may be applicable for several other diseases, especially those labeled 'complex diseases', and to environmental policy.

In Appendix B, I hypothesize that the evolutionary origins of sex determination, dioecy, and sex chromosomes were an ancillary consequence of cytosine methylation. I hypothesize that one cause for the relative fluidity of epigenetic signals over evolutionary time was a synergistic consequence of higher deamination rates and lower mismatch repair rates of methylated cytosine when compared with unmethylated cytosine. The keystone of this chapter is that cytosine methylation on a pair of autosomes drives Muller's ratchet and the subsequent degeneration of the Y or W chromosome. This is the first instance of anybody using the population genetic model of Muller's ratchet in

conjunction with changes in epigenetic signals. I explicate many testable predictions arising from this hypothesis and provide some evidence in corroboration.

In part, Appendix C is a special case of Appendix B, describing how sex determination may be due to cytosine methylation in dioecious lineages without noticeable sex chromosomes. I then describe how sex can then possibly be changed, possibly into a hermaphrodite. The goal of this work was to propose a method for creating a female individual for a species in which only males are still extant. Such a proposal also requires a methodology for regenerating differentiated organisms from undifferentiated tissue culture (callus), which constitutes the second part of this chapter. I propose that both sex change and tissue differentiation could be mediated by altering cytosine methylation levels. I am still trying to persuade botanic gardens and agricultural researchers to try such a protocol.

Appendix C was the genesis of all ideas in this dissertation and the first chapter that was written and published. In retrospect, Appendix C may appear unrefined and anomalously pragmatic, but it is the kernel of and/or motivation for all other chapters contained herein.

Appendix D is a mélange of two topics: the role of transposons in suppressing recombination (also discussed in Appendix B) and the evolutionary basis for meiosis. My hypothesis, (which I am developing more fully elsewhere; Gorelick, submitted-b is that resetting of epigenetic signals during meiosis and following syngamy is the crux of ploidy cycling in meiotic eukaryotes. Appendix D ostensibly focuses on ancient asexual animals and how they have managed to survive without outcrossing for tens of millions

of years. My hypothesis is that ancient asexual animal lineages are automictic hence reset their epigenetic signatures once or twice per generation and can thereby survive without outcrossing. I provide evidence from the literature on automixis or epigenetic resets for each of the known ancient asexual animal lineages.

Chapter 5, the conclusion, contains a wrap-up of the above work and a discussion of how this work can be improved and generalized. It is therefore, by necessity, the most speculative of the chapters. I try to indicate how far evolutionary epigenetic theory can carry us.

CHAPTER 2

FORMAL FISHERIAN DEFINITION OF EPIGENETICS AND EPISTASIS

ABSTRACT

The molecular revolution drove the neo-Darwinian paradigm towards considering nucleotides as the sole heritable component of chromosomes and thus nucleotides to be the sole source for heritable variation. However, epistatic and epigenetic interactions can cause phenotypic evolution with only changes in genotype frequency and epiallele frequency, respectively, *without any changes in allele frequency*. To resolve this conundrum and to suitably expand the neo-Darwinian paradigm, I built a multi-locus multi-allele formal Fisherian model of epistasis for diploid meiotic eukaryotes. The model presented here is restricted to diploid lineages with Mendelian inheritance, but without recombination and without within-locus dominance. I re-conceptualized epigenetics as a special form of epistasis by considering epialleles to be equivalent to very tightly linked enhancer alleles. As epistasis and epigenetics could provide large and previously hidden sources of heritable variation, this model enables us to study these questions both theoretically as well as through simulations. This work was not part of the dissertation, but is part of a future research project. Furthermore, this model proffers epistasis and epigenetics as reductionist mechanisms for producing seemingly holistic evolutionary changes, thereby partly bridging the schism between molecular biologists and quantitative geneticists.

INTRODUCTION

The modern paradigmatic definition of evolution is a change in allele frequency, with alleles almost invariably envisioned as nucleotide sequences, a paradigm that was largely driven by recent technological advances in gene sequencing. Unfortunately, this paradigm neglects the ubiquitous roles of epistasis and epigenetics, both of which can provide modes for evolution without any change in allele frequency (Wright, 1931; Waddington, 1940; Riggs, Martienssen & Russo, 1996a; Wolf, Brodie & Wade, 2000). Epistasis has largely been ignored because of the assumption that sufficiently powerful computational genomic tools will eventually account for the joint probabilistic effects between individual nucleotides. Epigenetics has largely been ignored because of the resilience of the sequence hypothesis and central dogma (Crick, 1970) and lack of quantification of epigenetic signals (Beck, Olek & Walter, 1999), e.g. lack of bisulfite sequencing (Frommer, McDonald, Millar, Collis, Watt, Grigg, Molloy & Paul, 1992).

Epistasis describes non-additive interactions between genes. If genotype frequencies change without any concomitant change in allele frequency, then epistatic interactions can result in phenotypic evolution (Goodnight, 1995). This provides a compelling reason to better understand the roles of complex (i.e. multi-locus) epistatic interactions, both theoretically as well as empirically. Epigenetics describes signals on

top of genetic signals. Epigenetic signals can be transmitted from parents to offspring, often almost as faithfully as genetic signals. From a classical Fisherian perspective, epigenetics can be thought of as a previously unrecognized form of epistasis. Therefore, my modeling effort focuses on epistasis, realizing that these models also apply to epigenetics. I develop the quantitative genetic complexities of epistasis, integrate modern molecular epigenetics within this epistatic model, and append these to the neo-Darwinian paradigm. In so doing, I remain within the confines of the traditional Fisherian approach of describing phenotypic effects of allelic substitutions.

EPIGENETIC SIGNALS ARE HERITABLE

Heritable portions of chromosomes are not only found in DNA, but also in chromosomal proteins and methylation. Methylation occurs as 5-methylcytosine on short DNA palindromes (usually as CpG dinucleotides) or as methylated lysine (and occasionally arginine) residues on histones (Ramsahoye, Biniszkiewicz, Lyko, Clark, Bird & Jaenisch, 2000; Jenuwein, 2001). Heterochromatic proteins other than histones also bind to DNA, with a predilection – but not a necessity – for binding to methylated cytosine (Miller, Schnedl, Allen & Erlanger, 1974; Miniou, Jeanpierre, Blanquet, Sibella, Bonneau, Herbelin, Fischer, Niveleau & Viegas-Péquignot, 1994; Wolffe & Matzke, 1999). These chromosomally bound proteins and methylation are collectively called epigenetic signals by molecular biologists and often also by developmental biologists. As an example, the epigenetic signals of genomic imprinting (i.e. differential methylation

between females and males) provide a molecular mechanism by which paternal or maternal inheritance can occur without having to invoke any nucleotide variation or loci on sex chromosomes. Environmental signals that alter the presence or absence of cytosine methylation or of heterochromatin provide facultative epigenetic signals that can cause evolution without any changes in nucleotide sequences (Holliday & Pugh, 1975). See chapters 4 and 5 for further examples. Because epigenetics have been so poignantly implicated in ontogeny of eukaryotes (Cowley & Atchley, 1992; Wolf et al., 2001), epigenetic changes without any changes in traditional allele frequency can potentially cause substantial phenotypic evolution. Definitive proof that epigenetic signals are heritable has surprisingly never been quantified using standard quantitative methods (with the molecular epigenetic signals serving as the phenotype, cf. Lewontin, 1992). However, there is compelling evidence that offspring have epigenetic signals closely resembling those of their parents. I devised quantitative genetic methods by which these epigenetic components of heritability can be statistically measured in experimental systems using classical macroscopic phenotypic data (see chapter 4). There is therefore an urgent need to more fully understand the roles of epistasis and epigenetics in quantitative genetics, in particular and evolutionary biology, in general.

Until recently, surprisingly little quantitative genetic work has been carried out on epistasis or epigenetics (Cheverud & Routman, 1995; Wolf et al., 2000; Wolf et al., 2001). Wright (1931) made evolutionary biologist acutely aware of epistasis, but he and subsequent workers confined attention to relatively simple mathematical models or verbal models. A formal mathematical definition of epistasis in terms of allelic substitutions at

one locus influencing the phenotypic effects of allelic substitution at another locus – i.e. so called physiological epistasis – has only ever been given for the two-locus two-allele case (Laubichler, 1997; Wagner, Laubichler & Bagheri-Chaichian, 1998). Richness is often gained by adding dimensions (Abbott, 1899). Extending the existing epistatic model to additional loci also forced me to further develop other important population genetic constructs, such as multi-locus linkage disequilibrium (chapter 3).

Even less quantitative genetic work has been carried out in epigenetics than in epistasis. Part of this is probably due to lack of a consistent operational definition of epigenetics (see chapters 1 and Appendix D). This is especially true for many cancer biologists who see no need for quantitative genetics being applied to epigenetic signals because, to them, epigenetic signals are only inherited via mitosis (see Appendix D). The few existing epigenetic models in quantitative genetic (Cowley & Atchley, 1992; Wolf et al., 2001) are of the statistical ilk, and do not explicitly model physiological epistasis.

DETAILS OF MODEL

For the case with no within locus dominance, I extended the existing formal Fisherian definition of epistasis from a two-locus two-allele model (Laubichler, 1997; Wagner et al., 1998) to a model that encompasses an arbitrary number of loci and alleles. Such a model can also extend to epigenetics by either making half of the loci into epiloci or, equivalently, by considering the existence of both a genetic allele and an epigenetic epiallele at each locus. For example, consider cytosine methylation (adenine can also be

methylated, but not in meiotic geneomes). Each nucleotide can be either methylated or not. This forms a binary epicode on top of the ternary nucleotide code, i.e. equal numbers of loci and epiloci. As another example, consider classical genes. Whether they are euchromatic or heterochromatic forms a binary code on top of the classical gene code, which will not necessarily be ternary. In either example, the epigenetic portion of the code acts by epistatically interacting with the genetic portion; cytosine methylation and heterochromatin generally act by suppressing gene activity. For the case with within locus dominance, I extended the existing two-locus two-allele model to two loci with an arbitrary number of loci. Throughout, I assume Mendelian inheritance and diploidy.

Wagner et al. (1998) give the following formulation for genotype values in the two-locus two-allele case. Let $G_{ij,kl}$ be the genotype value for the genotype with alleles *i* and *j* at the first diploid locus and alleles *k* and *l* at the first diploid locus, where *i*, *j*, *k*, and *l* can each only take on one of two values, conveniently labeled 1 and 2. They also let *A* and *B* designate the first and second loci, respectively. They then define several other terms as follows:

$$a_{A,kl} = \frac{1}{2}\left(G_{22,kl} - G_{11,kl}\right) \qquad e_{B\to A} = \frac{a_{A,22} - a_{A,11}}{2a_{A,11}}$$

$$E_{AB} = a_{A,11} \cdot e_{B\to A}$$

$G_{(AxA)12,kl}$ = additive plus additive-by-additive epistatic effects (i.e. breeding values)

$$d_{A,kl} = G_{12,kl} - G_{(AxA)12,kl} \qquad ed_{B\to A} = \frac{d_{A,22} - d_{A,11}}{2d_{A,11}}$$

$$edd_{12,12} = G_{12,12} - G_{(AxA)12,12} - d_{A,11}(1 + ed_{B\to A}) - d_{B,11}(1 + ed_{A\to B})$$

With these definitions, they completely describe all genotype values as follows:

$$G_{11,22} = G_{11,11} + 2a_{B,11} \qquad G_{22,11} = G_{11,11} + 2a_{A,11}$$

$$G_{22,22} = G_{11,11} + 2a_{A,11} + 2a_{B,11} + 4E_{AB}$$

$$G_{12,11} = G_{11,11} + a_{A,11} + d_{A,11} \qquad G_{11,12} = G_{11,11} + a_{B,11} + d_{B,11}$$

$$G_{12,22} = G_{11,11} + a_{A,11} + 2a_{B,11} + 2E_{AB} + d_{A,11}(1 + 2ed_{B\to A})$$

$$G_{22,12} = G_{11,11} + a_{B,11} + 2a_{A,11} + 2E_{AB} + d_{B,11}(1 + 2ed_{A\to B})$$

$$G_{12,12} = G_{11,11} + a_{A,11} + a_{B,11} + E_{AB} + d_{A,11}(1 + ed_{B\to A}) + d_{B,11}(1 + ed_{A\to B}) + edd_{12,12}$$

Before proceeding, I will slightly change notation. From herein out, let *i* and *j* represent loci (not alleles).

The first step in computing breeding values for physiological epistasis is to extend the enumeration of genotype values given above to multiple loci and alleles. Index the diploid genotypes at each locus. With $k(i)$ alleles at the i^{th} locus, there are $k(i)$ homozygotes and ½ $k(i)\cdot(k(i)\text{-}1)$ heterozygotes at that locus. Usually these homozygotes and heterozygotes are depicted in a two-dimensional punnett square array. To ease computation in subsequent steps, however, I re-index, converting the two-dimensional array of genotypes at a single locus (indexed by *f* and *m* for female and male homologues) to a one-dimensional array. Let $x(i)$ be the index, which merely is a way to count up all the elements in the punnett square.

$$x(i) = \left(k(i) - \tfrac{1}{2} f(i) + \tfrac{1}{2}\right) f(i) - \left(k(i) - m(i) + 1\right), \text{ with } f(i) \geq m(i).$$

With this indexing of diploid genotypes, homozygotes have indices that are evenly divisible by the number of alleles, k(i). In fact, define the functions

$$\theta(x(i)) = \begin{cases} 0 & if\ x(i) = 0 \\ 0 & if\ k(i) \mid x(i) \\ 1 & if\ k(i) \nmid x(i) \end{cases} = \begin{cases} 0 & if\ x(i)\ homozygous \\ 1 & if\ x(i)\ heterozygous \end{cases}$$

and

$$\phi(x(i)) = \begin{cases} 0 & if\ x(i) = 0 \\ 1 & if\ k(i) \nmid x(i) \\ 2 & if\ k(i) \mid x(i) \end{cases}$$

Also define a function (ξ) that equals zero for the first homozygote, one for heterozygotes that contain the first female allele (since we will have already accounted for additive-by-additive effects before computing dominance deviations), and two otherwise:

$$\xi(x(i)) = \begin{cases} 0 & if\ x(i) = 0 \\ 1 & if\ 1 \le x(i) \le k(i) - 1 \\ 2 & if\ x(i) \ge k(i) \end{cases}.$$

Then, we can completely describe all genotype values as follows:

$$G = G_0 + \left[E \cdot \prod_{i=1}^{n} \phi(x(i)) \right] + \sum_{i=1}^{n} \sum_{j \neq i} \left[\varphi(x(i)) a_{i,11} + \theta(x(i)) d_{i,11} \left(1 + \xi(x(i)) ed_{j \to i}\right) \right] + \left[edd \cdot \prod_{i=1}^{n} \theta(x(i)) \right]$$

or more precisely,

$$G = G_0 + \left[E \cdot \prod_{i=1}^{n} \phi(x(i)) \right] + \sum_{i=1}^{n} \sum_{j \neq i} \left[\varphi(x(i)) a_{k(i),11} + \theta(x(i)) d_{k(i),11} \left(1 + \xi(x(i)) ed_{k(j) \to k(i)}\right) \right] + \left[edd \cdot \prod_{i=1}^{n} \theta(x(i)) \right] \quad (1)$$

where $k(i)$ is one of the $k(i) - 1 = 1$ homozygous diploid genotypes at the i^{th} locus other than $x(k(i)) = 0$ and

$G = G[x(1), x(2)]$, $G_0 = G_{11,11} = G[x(1) = 0, x(2) = 0]$, and $E = E_{AB}$.

All that I have done in this paragraph is re-written the two-locus two-allele results of Wagner et al. (1998) in different notation; notation that can be extended to multiple loci and alleles.

To extend equation 1 to multiple alleles per locus, realize that wherever *i* and *j* (homozygous diploid genotypes at one of the two loci) appear as subscripts, we also now need another subscript that depends on which of the *k*(*i*)-1 or *k*(*j*)-1 homozygotes other than $x(k(i)) = 0$ we consider at that locus. We then have $\prod_{i=1}^{n}(k(i)-1)$ different versions of E_{AB}, albeit only one for each genotype $[x(1), x(2)]$. Thus for two-loci and multiple alleles, the results of Wagner et al. (1998) can be extended to

$$G[x(1)x(2)] = G[0,0] + \left[E[x(1), x(2)] \cdot \prod_{i=1}^{2} \phi(x(i)) \right]$$

$$+ \sum_{i=1}^{2} \sum_{j \neq i} \left[\varphi(x(i)) a_{k(i),x(j)=0} + \theta(x(i)) d_{k(i),x(j)=0} \left(1 + \xi(x(i)) ed_{k(j)\to k(i)}\right)\right] \qquad (2)$$

$$+ \left[edd \cdot \prod_{i=1}^{2} \theta(x(i)) \right]$$

which can be rewritten as

$$G[x(1)x(2)] = G[0,0] + \left[E[x(1), x(2)] \cdot \prod_{i=1}^{2} \phi(x(i)) \right]$$

$$+ \sum_{i=1}^{2} \sum_{j \neq i} \left[\varphi(x(i)) a_{k(i),x(j)=0} + \theta(x(i)) \left[d_{k(i),x(j)=0} \left(1 + \xi(x(i)) ed_{k(j)\to k(i)]}\right) + \theta(x(j)) edd \right]\right]$$

Extending equation 2 to multiple loci is much harder. We would first need to rename the terms d, ed, edd in Wagner et al. (1998) to $d^{(0)}$, $d^{(1)}, d^{(2)},\ldots,d^{(n)}$, where *n* is the number of loci, and extend the double summation to a multiple summation over all *n* loci. Here, for example, $d^{(4)}$, is the dominance deviation arising from the epistatic interaction between four heterozygous loci each with within locus dominance deviations. With a total of four loci, $d^{(4)}$ is the DxDxDxD epistasis coefficient. I have not yet managed to recursively define the higher order dominance deviations, $d^{(n)}$. However, things are quite a bit simpler if we ignore all within-locus dominance, in which case, we get

$$G[x(1)x(2)\cdots x(n)] = G[0,0,\ldots,0] + \left[E[x(1),x(2)\ldots x(n)]\cdot \prod_{i=1}^{n}\phi(x(i)) \right] + \sum_{i=1}^{n}\sum_{j\neq i}\left[\varphi(x(i))\cdot a_{k(i),x(j)=0}\right] \quad (4)$$

Here, there are $\prod_{i=1}^{n}(k(i)-1)$ distinct versions of E and of $a_{k(i),x(j)=0}$. Thus, using equation 4, all genotype values can be defined by $1+2\prod_{i=1}^{n}(k(i)-1)$ parameters, where the first parameter equals the genotype value $G[0,0,\ldots,0]$, of the genotype for which the first allele appears in the homozygous state at all loci.

To compute breeding values, we must combine genotype values with gamete frequencies, which are usually given by coefficients of linkage disequilibrium. Multi-locus linkage disequilibrium has never previously been defined for more than six loci, hence I have done this in chapter 3. Once we have the gamete frequencies and genotype

values for multiple loci and multiple alleles, the standard equations apply for computing average effects (Falconer & Mackay, 1996) and average irreducible effects (Laubichler, 1997), which give the additive genetic effects due to epistasis. From this, we can compute breeding values and consequently additive genetic variance *if and only if* we can compute recombination coefficients between multiple loci. This could possibly be done using a method akin to Lewontin (1964), although I have not done this. Without recombination, breeding value can simply be computed as the sum over the average effects of all alleles. Thus, the computations detailed herein only apply to Mendelian inheritance without recombination. My suspicion is that these two restrictions can be relaxed, but I have not yet been able to do this.

The most important result of the above computations is that it allows to asses the relative contribution of epistastic effects to additive genetic variance. This epistatic model also explicitly examines covariances between the functional units of selection, such as gametes and their alleles. If these covariances are negative, then traditional models in which no epistasis is assumed provide overestimates of the additive genetic variance, ultimately resulting in inflated estimates of heritability. The computations given above will eliminate that overestimate, thereby providing more accurate estimates of heritability if there exists any epistasis or epigenetics.

DISCUSSION

There are several substantial potential products from this modeling effort. First and foremost, this model will allow for prediction of the heritable variation due to epistasis and epigenetics. Historically, lack of sources for heritable variation has been a huge problem in evolutionary biology ever since Darwin. If epistasis or epigenetics provides a sizeable source for such variation, biologists will gain insight into the fuel that fires evolution. This could eliminate the evolutionary cul-de-sac predicted by Fisher's fundamental theorem of natural selection (Fisher, 1930), without having to invoke unrealistic levels of mutation.

This model places on firm theoretical ground evolution via environmental factors altering the faithful transmission of chromosomal signals, an idea that, in principle, even Darwin espoused (Darwin, 1872; Darwin, 1875). Although molecular evidence for such environmentally induced heritable changes was first identified over seventy-five years ago (position effect variegation; Sturtevant, 1925; Jollos, 1934), such environmentally induced evolution has never been incorporated into the neo-Darwinian paradigm. A firm theoretical understanding of epigenetics will allow us to more properly and accurately model those components of heritability that had earlier been ascribed to the statistical artifice of genetic-by-environmental interactions (GxE and norms of reaction) without any underlying molecular basis. See chapter 4 for details of how epigenetic effects may result in non-linear reaction norms through wholly linear mechanisms and the quantitative genetics of environmentally-induced heritable changes.

In quantitative genetics, neither epistasis nor epigenetics are usually thought of in terms of the mechanistic manner in which they affect allele substitutions. Epistasis is

usually thought of as a statistical deviation from additivity, i.e. genetic noise (Cheverud & Routman, 1995; Falconer & Mackay, 1996; Wade, Winther, Agrawal & Goodnight, 2001). Epigenetics is usually thought of as a mysterious developmental phenomenon (Cowley & Atchley, 1992; Schlichting & Pigliucci, 1998; Wolf et al., 2001), and not in terms of the mechanistic manner in which epiallelic substitutions affect allelic substitutions. The perspective that my proposed model effuses bridges this gap between formal quantitative genetics and molecular genetics. Furthermore, if epistasis or epigenetics contributes substantial additive variation (Wagner et al., 1998; Wade et al., 2001), then this perspective has the potential to bridge the enormous schism separating those who champion genomics and other highly reductionist notions and those few evolutionary researchers who have taken a decidedly holistic view of genetics and evolution (e.g. McClintock, 1950; Lewontin, 1991). Both reductionist and holistic evolutionary genetic viewpoints have much to offer and may be saying the same thing when examining epistasis or epigenetics.

CHAPTER 3

DECOMPOSING MULTI-LOCUS LINKAGE DISEQUILIBRIUM

ABSTRACT

We present a mathematically precise formulation of total linkage disequilibrium between multiple loci as the deviation from probabilistic independence and provide explicit formulae for all higher order terms of linkage disequilibrium, thereby combining Dausset *et al*.'s (1978) definition of linkage disequilibrium with Geiringer's (1944) approach. We recursively decompose higher order linkage disequilibrium terms into lower order ones. Our greatest simplification comes from defining linkage disequilibrium at a single locus as allele frequency at that locus. At each level, decomposition of linkage disequilibrium is mathematically equivalent to number theoretic compositions of positive integers, i.e. we have converted a genetic decomposition into a mathematical decomposition.
[This chapter is published in *Genetics*; see Appendix E. This was also the only chapter co-authored (Gorelick and Laubichler) hence I have used the pronoun "we" throughout.]

INTRODUCTION

A precise measurement of linkage disequilibrium is required for studying virtually any phenomenon in multi-locus population genetics. This is especially true for explicit multi-locus models that investigate the contributions of physiological epistasis to additive genetic variance (Cheverud & Routman, 1995; Wagner et al., 1998; Wagner &

Laubichler, 2000). Linkage disequilibrium is usually defined as the deviation from probabilistic independence between alleles at two different loci. This deviation from independence can have different causes, such as a lack of independent segregation or recombination, or any number of other evolutionary forces. The presence of linkage disequilibrium (gametic disequilibrium) is thus an indication that either stochastic (e.g. drift) or deterministic (e.g. selection, gene flow) evolutionary forces have been acting on a population (Hedrick, 2000; Ardlie, Kruglyak & Seielstad, 2002).

The classical definition of linkage disequilibrium, D, follows the probability theory definition of deviation from independence. Independence of two events B and C means that $\Pr(BC) = \Pr(B) \cdot \Pr(C)$, where Pr is probability and BC is the joint distribution of B and C, so that the deviation from independence is measured as $D = \Pr(BC) - \Pr(B) \cdot \Pr(C)$. Changing notation slightly to let $A_{k(i)}$ designate the k^{th} allele at the i^{th} locus gives the linkage disequilibrium between the alleles at two loci, D_2 as $D_2 = \Pr(A_{k(1)}A_{k(2)}) - \Pr(A_{k(1)}) \cdot \Pr(A_{k(2)})$, where Pr represents probability and $A_{k(1)}A_{k(2)}$ represents the joint occurrence of $A_{k(1)}$ and $A_{k(2)}$ in a single haploid gamete. In most modern interpretations of probability theory, the primitive concept of 'probability' is interpreted as a relative frequency, therefore $\Pr(A_{k(1)})$ is the same as the frequency of allele k at locus 1.

The quintessential examples of linkage disequilibrium are coadapted gene complexes, in which several loci are tightly linked because they provide a large selective advantage if they occur together. In these cases, linkage disequilibrium is maintained by

selection. Although coadapted gene complexes are implicit in Wright's shifting balance hypothesis (Wright, 1931), have been used to explain outbreeding depression (Dobzhansky, 1948; Lynch, 1991), and are frequently cited as evolutionary hypotheses (Palopoli & Wu, 1996; Rawson & Burton, 2002), the linkage disequilibrium of these purported coadapted gene complexes is almost never quantified. This is particularly surprising given the well-cited paper by Geiringer (1944), in which she provides most of the algorithm for computing higher order linkage disequilibrium coefficients. In this paper, we complete and simplify Geiringer's formulation and then show how the sums of products of those coefficients equals the definition of (total) linkage disequilibrium as the deviation from probabilistic independence given by Dausset *et al*. (1978).

Methodologically, we follow Geiringer's lead and decompose higher order linkage disequilibrium into lower order linkage disequilibrium terms. In other words, we take a top-down approach to defining multi-locus linkage disequilibrium, rather than the bottom-up approach followed by virtually everyone since Geiringer (1944). Lewontin (1974) is typical of the bottom-up approach. There are very few other top-down decomposition approaches in population genetics such as Bulmer's (1980) decomposition of multi-locus epistasis or Wagner and Laubichler's (2000) character decomposition approach.

In this paper, we first define linkage disequilibrium at a single locus as the allele frequency at this locus, which will greatly simplify notation. Second, we extend the definition of linkage disequilibrium to multiple loci by invoking compositions of positive integers. Our decomposition of multi-locus linkage disequilibrium is entirely consistent

with the standard definitions for two loci, as well as its previous extensions to three, four, and six loci (Geiringer, 1944; Bennett, 1954; Hastings, 1984). Third, we show how this definition is entirely consistent with the notion of linkage disequilibrium as the deviation from probabilistic independence.

DECOMPOSITION OF MULTI-LOCUS LINKAGE DISEQUILIBRIUM

Define the one locus coefficient of linkage disequilibrium, D_1, as $D_1\left(A_{k(i)}\right) = \Pr\left(A_{k(i)}\right)$. This definition may appear paradoxical, but it dramatically simplifies notation for the decomposition of multi-locus linkage disequilibrium. In elementary algebra we have the analogous problem of defining the algebraic expression x^n when n equals zero (Lakoff & Núñez, 2000). Note that our definition of a locus encompasses protein coding loci, QTLs, and even single nucleotides.

The formulas for two and three locus multi-locus linkage disequilibrium, in which $D_1\left(A_{k(i)}\right)$ was substituted for $\Pr\left(A_{k(i)}\right)$, are defined by Hastings (1984) as:

$$D_2 = \Pr\left(A_{k(1)}A_{k(2)}\right) - D_1\left(A_{k(1)}\right)\cdot D_1\left(A_{k(2)}\right)$$

$$\begin{aligned} D_3 = \Pr\left(A_{k(1)}A_{k(2)}A_{k(3)}\right) &- D_1\left(A_{k(1)}\right)\cdot D_1\left(A_{k(2)}\right)\cdot D_1\left(A_{k(3)}\right) \\ &- D_1\left(A_{k(1)}\right)\cdot D_2\left(A_{k(2)}A_{k(3)}\right) \\ &- D_1\left(A_{k(2)}\right)\cdot D_2\left(A_{k(1)}A_{k(3)}\right) \\ &- D_1\left(A_{k(3)}\right)\cdot D_2\left(A_{k(1)}A_{k(2)}\right) \end{aligned}$$

The pattern here is that D_n equals $\Pr\left(A_{k(1)}A_{k(2)}\ldots A_{k(n)}\right)$ minus all possible products of lower order linkage disequilibrium coefficients, such that each term has all of its subscripts adding up to *n*. The key to writing down an explicit formula for D_n is that the phrase "all possibilities of the subscripts adding up to *n*" refers to partitions of the positive integer *n* (Andrews, 1976). A partition π of a positive integer n is a set of positive integers that add up to n, i.e. $\pi = \{n_1, n_2, \ldots, n_m\}$ such that $\sum_{i=1}^{m} n_i = n$. The set of all partitions of n is designated $p(n)$, e.g. $p(5) = \{\{5\}, \{4,1\}, \{3,2\}, \{2,2,1\}, \{3,1,1\}, \{2,1,1,1\}, \{1,1,1,1,1\}\}$. To define multi-locus linkage disequilibrium, we have to add over all partitions, excluding the trivial partition $\pi = \{n\}$ and permute over all alleles for a given number of loci. However, the order of elements of the partition matter, hence we construct the number-theoretic compositions *c* of the positive integer *n* (Andrews, 1976). For example, all of the compositions for the partition $\pi = \{2,2,1\}$ are the ordered triples $(2,2,1)$, $(2,1,2)$, and $(1,2,2)$. Using these mathematical notions we can generalize the two and three locus cases to define linkage disequilibrium of n loci as follows:

$$\mathrm{D}_n\left(A_{k(1)}, A_{k(2)},\ldots A_{k(n)}\right) = \Pr\left(A_{k(1)}A_{k(2)}\ldots A_{k(n)}\right) - \sum_{\substack{\text{all compositions } c \text{ of } n \\ \text{except } c=(n)}} \left[\prod_{n_i \in c} D_{n_i}(\cdots)\right] \tag{1a}$$

where $n_i \in c$ means that $n_i \in c$ is a scalar component of the vector c. Equivalently,

$$D_n\left(A_{k(1)}A_{k(2)}...A_{k(n)}\right)=\Pr\left(A_{k(1)}A_{k(2)}...A_{k(n)}\right)-\sum_{\substack{\sum_{i=1}^{m} n_i=n \\ 1\le n_i<n \\ 1\le m\le n}}\left[\prod_{i=1}^{m} D_{n_i}\left(\cdots\right)\right]. \quad (1b)$$

There is only way to decompose n into a single positive integer: $c=(n)$.

Therefore we can also write the highest order coefficient of linkage disequilibrium as

$D_n\left(A_{k(1)}, A_{k(2)},...A_{k(n)}\right)=\sum_{c=(n)}\left[\prod_{n_i\in c} D_{n_i}\left(\cdots\right)\right]$, where the summation only has a single term and the product only has a single factor. Therefore, equation 1a yields

$$\Pr\left(A_{k(1)}, A_{k(2)},...A_{k(n)}\right)=\sum_{\text{all compositions } c \text{ of } n}\left[\prod_{n_i\in c} D_{n_i}\left(\cdots\right)\right], \quad (2)$$

which we will use below.

Equation 1 has never been written explicitly for multi-locus linkage disequilibrium, even though special cases have been given by Geiringer, 1944, Bennett, 1954, and Hastings, 1984). The only explicit definition previously given for multi-locus linkage disequilibrium is the following due to Dausset, et al. (1978):

$$\mathbf{D}_n\left(A_{k(1)}A_{k(2)}...A_{k(n)}\right)=\Pr\left(A_{k(1)}A_{k(2)}A_{k(3)}...A_{k(n)}\right)-\prod_{i=1}^{n} D_1\left(A_{k(i)}\right) \quad (3)$$

which we call total linkage disequilibrium, $\mathbf{D}_n$,where we have again replaced $\Pr\left(A_{k(i)}\right)$ with $D_1\left(A_{k(i)}\right)$. We refer to $\mathbf{D}_n$ as total linkage disequilibrium because, as we show below, all of the non-bolded linkage disequilibrium coefficients

$D_1, D_2, D_3, \ldots, D_n$ can be independent from one another and contribute to $\mathbf{D}_n$.

Equation 3 has a simple heuristic interpretation: $D_n\left(A_{k(1)}...A_{k(n)}\right)$ measures how far the haploid genotype at all n loci deviates from probabilistic independence.

We are now ready to derive the relationship between D_n and $\mathbf{D}_n$. In equation 3, substitute $\sum_{\text{all compositions } c \text{ of } n} \left[\prod_{n_i \in c} D_{n_i}(\cdots)\right]$ for $Pr\left(A_{k(1)},...A_{k(n)}\right)$ (see equation 2),

yielding

$$\mathbf{D}_n\left(A_{k(1)}, A_{k(2)},...A_{k(n)}\right) = \sum_{\text{all compositions } c \text{ of } n} \left[\prod_{n_i \in c} D_{n_i}(\cdots)\right] - \prod_{i=1}^{n} D_1\left(A_{k(i)}\right)$$

. The last term in this equation is simply the value of $\prod_{n_i \in c} D_{n_i}(\cdots)$ for the composition $c = (1,1,1,\ldots,1)$, i.e. $n = \underbrace{1+1+\ldots+1}_{n \text{ times}}$. Therefore, combining equations 3 becomes

$$\mathbf{D}_n\left(A_{k(1)}, A_{k(2)},...A_{k(n)}\right) = \sum_{\substack{\text{all compositions } c \text{ of } n \\ \text{except } c=(1,1,1,\ldots,1)}} \left[\prod_{n_i \in c} D_{n_i}(\cdots)\right] \qquad (4a)$$

or, equivalently,

$$\mathbf{D}_n\left(A_{k(1)}, A_{k(2)},...A_{k(n)}\right) = \sum_{\substack{\sum_{i=1}^{m} n_i = n \\ 1 \le n_i < n \\ 1 \le m < n}} \left[\prod_{i=1}^{m} D_{n_i}(\cdots)\right] \qquad (4b)$$

Equation 4 provides the crucial link between deviations from independence ($\mathbf{D}_n$) and the linkage disequilibrium coefficients D_n computed by Geiringer (1944) and her intellectual successors by decomposing $\mathbf{D}_n$ into the terms D_{n_i}, where $\sum n_i = n$.

DISCUSSION

We have converted the genetics problem of decomposing linkage disequilibrium into the mathematical problem of decomposing positive integers into their additive parts, all while maintaining the convenient heuristic definition of total linkage disequilibrium as the deviation from independence. Unlike Geiringer (1944), we can write down an explicit formula for multi-locus linkage disequilibrium because we invoke partitions of integers and define $D_1(A) = \Pr(A)$, thereby merging her notion of linkage disequilibrium with those of Dausset et al. (1978).

One immediate consequence of our decomposition approach is that the single highest-order coefficient of linkage disequilibrium, D_n, cannot be examined in isolation. Because $D_n\left(A_{k(1)}, A_{k(2)}, \ldots A_{k(n)}\right) = \sum_{\substack{\text{all compositions } c \text{ of } n \\ \text{except } c=(1,1,1,\ldots,1)}} \left[\prod_{n_i \in c} D_{n_i}(\cdots) \right]$, we need to examine all lower order linkage disequilibrium coefficients, $D_{n_i}(\cdots)$ with $n_i < n$. All of the subscripted linkage disequilibrium coefficients

$D_1, D_2, D_3, \ldots, D_n$ can be independent from one another and all contribute to $\mathbf{D}_n$, which we therefore call total linkage disequilibrium.

Multi-locus definitions of linkage disequilibrium have not been used very often in empirical studies because of the large number of inputs and linkage disequilibrium coefficients that must be analyzed ($2^n - 1$). Currently, even third-order linkage disequilibrium is seldom measured (Thomson & Baur, 1984). However, explicit terms for multi-locus linkage disequilibrium are of theoretical importance.

Multi-locus definitions of linkage disequilibrium have not been used very often in empirical studies because of the large number of inputs and linkage disequilibrium coefficients that must be analyzed ($2^n - 1$ of each), but are of theoretical importance. Currently, even third-order linkage disequilibrium is seldom measured (Thomson & Baur, 1984).

One important theoretical application is the analysis of multi-locus epistasis. Cheverud and Routman (1995) developed a model two-locus for physiological epistasis that has been further refined by Wagner et al. (1998). In order analyze the evolutionary consequences of epistasis in these models, one has to first define linkage disequilibrium for a subset of the loci. Thus, to extend models of physiological epistasis to multiple loci, we must first define linkage disequilibrium for that subset of loci, which we have just done. Models of multi-locus epistasis will be crucial in debates over what factors maintain coadapted gene complexes, increase additive genetic variance, and foster speciation (Goodnight, 1988; Goodnight, 1995; Wade & Goodnight, 1998).

CHAPTER 4

ENVIRONMENTALLY-ALTERABLE ADDITIVE GENETIC EFFECTS

ABSTRACT

Environment in the parental generation can cause heritable changes in subsequent generations. In part, these heritable changes exist in the form of environmentally-alterable epigenetic signals of cytosine methylation and chromatin formation, although it is still an open empirical question how prevalent and how heritable these effects are. Parent-offspring regression models can be used to estimate components of additive genetic variance that are caused by environmentally-alterable signals, which is a generalization of indirect additive genetic variance. If environmentally-alterable additive genetic variance is large compared with direct additive genetic variance, then this confounds prediction of evolutionary trajectories, but (i) provides a mechanism by which environmental variance directly increases additive genetic variance, (ii) implies that environmental variance can cause evolutionary novelty, (iii) provides a mechanism underlying phenotypic plasticity, and (iv) and may provide an explanation for why plants are more phenotypically plastic than animals. Under the assumptions of no interaction variance (neither epistasis nor genotype-by-environment interaction variance), phenotypic variance can be decomposed into components due to nature and nurture if and only if environmentally-alterable additive genetic variance is zero.

ENVIRONMENTALLY-ALTERABLE EPIGENETIC SIGNATURES

How does ecology impinge on evolution? The simplest and most common answer is via selection, including notions of ecological speciation (Schluter, 1998). But quantitative genetics has also been extended to include other ecological influences, such as genotype-by-environment interactions, phenotypic plasticity, and maternal effects (Falconer & Mackay, 1996). In each of these instances, it is the environment of the offspring that affects the mapping from its genotype to phenotype (Lewontin, 1992). Here, I extend these notions to environment of parents heritably affecting the genotypes (and hence phenotypes) of their descendants.

Environmental perturbations can heritably alter organisms and are most likely to do so by heritably altering epigenetic signals. I use the term epigenetic here to refer to molecules that reside on top of DNA nucleotides ("epi-" is Greek for "on top of"), and not in the sense of developmental signals. There are, however, very large areas of overlap between these two definitions of epigenetic, e.g. the model herein is very close to Waddington's (1957) notion of environmentally-induced gene mutations (also see Holliday, 1994). It is also important to note that by environment, I am referring to that aspect of the environment that is measured (e.g. Bulmer, 1980) and not simply the residual from a regression or analysis of variance (e.g. Fisher, 1930).

There are several examples of environmentally-alterable epigenetic signals that are heritable. Parental behavior and stress responsivity in rats can be due to stress levels heritably altering cytosine methylation signatures (Champagne & Meaney, 2001; Meaney, 2001; Weaver, Szyf & Meaney, 2002). Coat color in mice can be heritably

altered by feeding the parents varying levels of methylation (Wolff, Kodell, Moore & Cooney, 1998; Rakyan & Whitelaw, 2003). Heritable changes in cytosine methylation have been implicated in shade avoidance in plants (Tatra, Miranda, Chinnappa & Reid, 2000). Canalization of genetic sex determination from an ancestral state of temperature-dependent sex determination in several lineages of amphibians, reptiles, and plants may be due to inheritance of altered epigenetic signals (Dorazi, Chesnel & Dournan, 1995; Gorelick & Osborne, 2002; Gorelick, 2003; Prahald, Pilgrim & Goodwin, 2003). Environmentally-induced heritable changes have also been induced in lineages that largely lack cytosine methylation, and instead utilize chromatin formation, including histone acetylation, for regulation, such as *Drosophila,* the nematode *Caenorhabditis elegans,* and the fission yeast *Schizosaccharomyces pombe* (Jollos, 1934; Ekwall, Olsson, Turner, Cranston & Allshire, 1997; Wolffe & Matzke, 1999; Gowher, Leismann & Jeltsch, 2000; Lyko, Ramsahoye & Jaenisch, 2000).

For environmentally-alterable epigenetic effects to be of evolutionary importance, heritable epigenetic signals must be transmitted across multiple successive generations, and not just from parents to offspring. Epigenetic effects that only get transmitted across one pair of generations can instead be modeled as special environment effects. Cytosine methylation and chromatin formation seem to be heritable over many successive generations, although few studies have been done to corroborate this. Grandparental inheritance of genomic imprinting has been documented (Croteau, Andrade, Huang, Greenwood, Morgan & Naumova, 2002). Tenuous evidence of these epigenetic signals being transmitted over many generations is circumstantial, such as inheritance of

genomic imprinting diseases (although these could be due to classical DNA mutations, such as a heritable defect in genes coding for methyltransferases; Ehrlich, 2000). The most convincing evidence of epigenetic inheritance across many generations comes from chromosomal banding. Because G-banding is consistent within a lineage over many generations, it has great utility in systematics and phylogenetics (Solari, 1994).

Heuristically, it makes sense that epigenetic signals should be heritable, albeit less so than DNA nucleotides. Epigenetic signals that reside on top of DNA nucleotides are analogous to an electrical wire, with DNA nucleotides forming the copper conducting portion, the epigenetic signals forming a heterogeneous (variegated) layer of insulation, and the entire genome forming a bundled cable of wires/chromosomes. The insulation can only act by altering the electrical and magnetic signals flowing through the wire (gene regulation), especially if the insulation is thin in places or if there is shorting and cross talk between individual wires in the bundled cable. Furthermore, it is much easier to damage the insulation (epigenetic signatures), say with a knife or heat, than it is to damage the underlying copper conducting wire (DNA nucleotides). Some of this environmental damage to epigenetic signals is corrected during meiosis and syngamy, but many of the epigenetic changes are transmitted to the next generation. Another way that this analogy works is that as bundled cables age, the insulation around the wires changes (degrades) in a fairly predicable way, e.g. telomere degradation with age, whereas the copper conducting portion of the wire seems relatively unaffected.

Environmentally-alterable heritable effects other than epigenetic signals also exist. For example, in most animals and plants, environmental condition of maternal

parents affects the nutritional provisioning that they provide their embryos, thereby affecting their offspring's phenotype. Therefore, although I have focused on environmentally-alterable heritable epigenetic effects, other forms of environmentally-alterable additive genetic effects exist, such as maternal effects (Kirkpatrick & Lande, 1989). One of the remarkable aspects of the early quantitative genetic work of Fisher (1930), Wright (1931), and Lush (1937) is that they developed quantitative genetics before biologists realized that DNA is largely contained in chromosomes (Caspersson & Schultz, 1940 citing Brachet, 1940). Thus, the simple quantitative genetic model that I present below can be used to model any form of environmentally-alterable heritable effect, regardless of underlying molecular mechanism.

The environmentally-alterable additive genetic effects introduced herein are generalizations of maternal environmental effects, $me^{*}_{m(t-1)}$, introduced by Wolf and Brodie (1998). Their maternal environmental effects are the environment *created by* the maternal parent, which they show are equivalent to a non-additive G x E interaction effect. By contrast, I consider the environmental condition, $c(t)$, that the maternal parent was herself *subjected to* and show that this is an additive effect (if I were using Wolf and Brodie's labels for time, I would re-label, $c(t)$ as $c(t-1)$, but have chosen not do so). Using an engineering metaphor, Wolf and Brodie (1998) consider $me^{*}_{m(t-1)}$ to be the environmental *output* of the parental phenotype, whereas I consider $c(t)$ to be the environmental *input* to the parental phenotype.

This paper is a simple extension of existing neo-Darwinian methods and is not meant as a replacement of theory that has held up very well for three-quarters of a century.

QUANTITATIVE GENETIC MODEL

I use a simple linear regression model in which the dependent variable is phenotype of the offspring, $z(t+1)$, and the independent variables are additive genetic effects of offspring, $a(t+1)$, stochastic environment of the offspring, $e(t+1)$, and measured or macro-environment of the offspring, $c(t+1)$. I follow Bulmer's (1980) convention of distinguishing the stochastic environment, *e*, from the measured environment or condition, *c*. Total environmental variance is therefore the sum of stochastic and measured environmental variances. Following Kirkpatrick and Lande (1989), *t* represents the parental generation and *t*+1 the offspring generation. To keep this paper conceptually simple, I have not explicitly included any epistatic effects, but have included G x E interaction terms. Because the additive genetic effects of offspring, $a(t+1)$, are not directly observable, invoke the almost universally accepted assumption that phenotype of the parents, $z(t)$, can be used as its proxy (Lush, 1940; Lynch & Walsh, 1998: 48-49). Begin with the following standard parent-offspring regression model:

$$z_{ij}(t+1) = G\, a_i(t+1) \qquad \text{(direct additive genetic variance)}$$

$$
\begin{aligned}
&+ K\, c_j(t+1) && \text{(measured environmental variance)} \\
&+ e_i(t+1) && \text{(stochastic environmental variance)} \quad (1) \\
&+ I\, z_i(t+1) \cdot c_j(t+1) && \text{(G x } E_{\text{offspring}} \text{ interaction variance)} \\
&+ b(t+1) && \text{(regression bias)}
\end{aligned}
$$

where i designates families and j designates conditions. G is the matrix of additive genetic variances, while K measures the measured (macro-) environmental influences. I use K instead of C only because upper and lower case C are difficult to distinguish in most typefaces. If no environmental variables are measured, simply delete K and c from equation 1. I is the G x E interaction matrix.

Examining variances in the scalar case of Equation 1, yields the well-known quantitative genetic variance decomposition, $V_P=V_A+(V_C+V_E)+V_{GxE}$, where V_P is phenotypic variance, V_C and V_E are the environmental variance (non-stochastic and stochastic parts, respectively), and V_{GxE} is the interaction variance between the offspring's genotype and its measured environment. The nature versus nurture dichotomy arises naturally from this variance decomposition, especially when V_{GxE} is zero: V_A is nature and (V_C+V_E) is nurture.

If measured parental environment affects offspring phenotype, then it should be added to the parent-offspring regression. With this new time-lagged independent variable of measured parental environment, $c_j(t)$, we must add two new terms to equation 1, yielding:

$$
\begin{aligned}
z_{ij}(t+1) = G\,a_i(t+1) \quad & \text{(direct additive genetic variance)} \\
+ H\,c_j(t) \quad & \text{(environmentally-alterable additive variance)} \\
+ K\,c_j(t+1) \quad & \text{(measured environmental variance)} \\
+ e_i(t+1) \quad & \text{(stochastic environmental variance)} \qquad (2) \\
+ I\,z_i(t+1)\cdot c_j(t+1) \quad & \text{(G x } E_{\text{offspring}} \text{ interaction variance)} \\
+ J\,z_i(t+1)\cdot c_j(t) \quad & \text{(G x } E_{\text{parent}} \text{ interaction variance)} \\
+ b(t+1) \quad & \text{(regression bias)}
\end{aligned}
$$

Call the variance of *H*, V_{EA}, the environmentally-alterable additive genetic variance because it depends on the measured parental environment. V_{EA} is still part of additive genetic variance because it depends upon the parents. Environmentally-alterable additive genetic variance, *H*, appears to be a new entity in quantitative genetics. *J* and its variance $V_{GxE\text{-}parent}$ are measures of the interaction between offspring environment and *parental* condition. The scalar case of the variance decomposition arising from Equation 2 is $V_P = (V_A+V_{EA}) + (V_C+V_E) + (V_{GxE}+V_{GxE\text{-}parent})$, where (V_A+V_{EA}) is the additive component of variance, (V_C+V_E) is the environmental component of variance, and $(V_{GxE}+V_{GxE\text{-}parent})$ is the genotype-by-environment interaction component of variance. Actually, because *H* and *J* have been omitted independent variables from all previous parent-offspring regression models, those previous models may have produced seemingly anomalous and non-linear behaviors. A regression model has to allocate those

effects to some independent variable that has been included in the model (see the section on phenotypic plasticity, below).

H and *J* (V_{EA} and $V_{GxE\text{-}parent}$) are generalizations of indirect genetic effects (Moore, Brodie & Wolf, 1997; Wolf, Brodie, Cheverud, Moore & Wade, 1998). Indirect genetic effects on an individual are those due to the phenotype of other individuals. Thus, maternal effects are a special form of indirect genetic effects. *H* and *J* are coefficients on parental condition, $c_j(t)$. Nonetheless, it is worthwhile singling out this special case of indirect genetic effects because existing mathematical models only include *condition of the focal individual*, and not *condition of conspecific individuals*, despite including phenotypes of all individuals. The only place that condition of non-focal individual seems to appear in the literature is the path diagram of Box 2c in Wolf et al. (1998).

PHENOTYPIC PLASTICITY ARISING AS OMITTED VARIABLE BIAS

What if environmental condition of the parents or more distant relatives should be included in the estimation of heritability, but had previously been omitted? Then $c_j(t)$ becomes an omitted variable from the regression, and there exists a formula for estimating the bias from such omissions (Greene, 1997). In particular,

$$E(bias) = G + H + \mathrm{cov}\left(a_i(t+1), c_j(t)\right) + \mathrm{cov}\left(\begin{pmatrix} a_i(t+1) \\ \cdots \\ c_j(t+1) \end{pmatrix}, c_j(t)\right) \cdot \mathrm{var}\begin{pmatrix} a_i(t+1) \\ \cdots \\ c_j(t+1) \end{pmatrix}^{-1}, \text{ where an}$$

ellipses designates concatenation of two vectors. If we assume that $c_j(t) \approx c_j(t+1)$ and $\text{cov}\left(a_i(t+1), c_j(t)\right) = 0$, then $E(bias) \approx G + H + \text{cov}\left(a_i(t+1), c_j(t)\right) \cdot \text{var}(a_i(t+1))$. For scalar values, the second term in this last equation is approximately equal to $V_{G\,x\,H} \approx V_{G\,x\,K}$. V_K is usually known as V_E, thus $E(bias) \approx V_G + V_E + V_{G\,x\,E} = V_E + V_{PL}$, where $V_{PL} = V_G + V_{G\,x\,E}$ is Scheiner and Goodnight's (1984) measure of phenotypic plasticity for linear and non-linear reaction norms. Thus, inclusion of the parent's condition in the linear parent-offspring regression of equation 2 accounts for phenotypic plasticity of reaction norms.

Notice, however, that $V_{PL} = V_G + V_{G\,x\,E}$ only provides a good measure of the spuriously non-linear reaction norms if $c_j(t) \approx c_j(t+1)$. What makes this result peculiar is that $c_j(t) \approx c_j(t+1)$ implies that the input data to the regression will be highly multi-collinear, thereby grossly inflating the standard error of estimate (Greene, 1997). If $c_j(t) \neq c_j(t+1)$ or $\text{cov}\left(a_i(t+1), c_j(t)\right) \neq 0$, then we must incorporate both direct and environmentally-alterable components of additive genetic variance into a single vector called $a_{ij}^*(t+1) = \begin{pmatrix} a_i(t+1) \\ \cdots \\ c_i(t) \end{pmatrix}$ whose covariance is

$G^* = \begin{pmatrix} G & \text{cov}\left(a_i(t+1), c_j(t)\right) \\ \text{cov}\left(a_i(t+1), c_j(t)\right) & H \end{pmatrix}$. Then equation 2 can be rewritten as

$z_{ij}^*(t+1) = G^* a_{ij}^*(t+1) + e_i(t+1) + b(t+1)$. If macro-environmental condition is truly exogenous, then $\text{cov}\left(a_i(t+1), c_j(t)\right) = 0$ and the above equation reduces to equation 2.

This analysis indicates that there is not good theoretical justification for using $V_{PL} = V_G + V_{G\,x\,E}$ as a simple scalar measure of phenotypic plasticity. Instead *H* and *J* (i.e. V_{EA} and $V_{GxE\text{-}parent}$) provide theoretically sound measures of these putative non-linearities. I say putative because, if there are no G x E interactions terms *I* and *J*, then *H* is a coefficient from a purely linear regression model.

DISCUSSION

Critics might argue that environmentally-alterable effects are simply another form of mutation and should therefore be incorporated into existing quantitative genetic models. Such an approach, however, ignores the correlation between two important independent variables that are already part of those very models. The direct additive genetic effects, $a_i(t+1)$, can be dependent upon parental measured environment, $c_j(t+1)$, because environmental perturbations can alter the genotype (epigenotype) in systematic ways. Ignoring this correlation is tantamount to ignoring a possibly important underlying mechanism…one that we were trying to model in the first place!

Consider equation 1 in which there are no environmentally-alterable additive genetic effects. With the standard assumption that *G* is constant over time, phenotype of one generation can be used to predict the phenotypes of subsequent generations using the standard formula $\Delta\bar{z} = GP^{-1}\beta$, where *P* is phenotypic variance, β is the selection coefficient, and this equation can be iterated over successive generations. By contrast, if

environmentally-alterable additive genetic variance for a trait (or suite of traits) is large compared with direct additive genetic variance, then this decimates the predictive power of this quantitative genetic model. With environmentally-alterable additive genetic variance as modelled by equations 2 and 3, $\Delta\bar{z}$ is a function of $c(t)$. Thus, with environmentally-alterable additive genetic variance, the only way to predict the evolutionary trajectory is to know the time history of measured environmental condition.

The fact that standard quantitative genetic models such as equation 1 have worked so well for many decades indicates that environmentally-alterable additive genetic effects are either not that common or not that large. Nonetheless, the fact that standard quantitative genetic models sometimes do not seem valid indicates that we should in some instances consider environmentally-alterable additive genetic effects.

Although presence of environmentally-alterable additive genetic variance confounds quantitative genetic prediction, it also provides a simple route by which exogenous spatial or temporal environmental variation injects additive genetic variation into a population. In this way, environmental variance can cause additive genetic variance even in populations for which additive genetic variance was originally zero. Environmentally-alterable additive genetic effects can be a source of evolutionary novelty. This could play a role in such phenomena as El Niño Southern Oscillations and glacial/interglacial cycles.

Environmentally-alterable heritable epigenetic signals provide a molecular mechanism that may underlie G x E interactions. If the environmentally-altered epigenetic signal is reset during meiosis or syngamy, then this results in a standard G x

$E_{offspring}$ interaction. If the environmentally-altered epigenetic signal is transmitted to offspring, then this results in a G x E_{parent} interaction. In either instance, G x E interactions are still non-linearities in an otherwise linear model, but epigenetic theory allows G x E interactions to be something other than mere statistical artifacts.

By explicitly including parental environment in parent-offspring regression models, I have explicated an important generalization of indirect genetic effects. Parental environment is different from and independent of the offspring environment that is due to indirect genetic effects, such as the offspring environment created by parents or other conspecifics (Moore et al., 1997; Wolf & Brodie, 1998; Wolf, 2003). Equation 2 allows us to estimate how exogenous parental environment affects evolutionary trajectories. By knowing the time history of parental environmental conditions, we can estimate the response of a population to selection. Additive genetic variance decomposes into an endogenous piece describing classical direct and indirect additive genetic effects (where the indirect piece includes classical maternal and paternal effects) and an exogenous piece describing environmentally-alterable additive genetic variance. Environmentally-alterable additive genetic variance, V_{EA}, provides a linear measure of what were probably often previously referred to as non-linear reaction norms. I have also introduced a truly non-linear component of phenotypic plasticity, *J*, which measures the interaction between offspring genetics and parental environment. This model provides a theoretically sound measure of phenotypic plasticity and does so by positing a way that environmental variance gets converted to additive genetic variance. This model also provides a quantitative genetic model by which to estimate the additive genetic

component of phenomena that are known to be heritably altered by environmental perturbations, such as maternal provisioning of embryo nutrients or the environmentally-alterable epigenetic signals of cytosine methylation.

Focusing on cytosine methylation for a moment, environmentally-alterable additive genetic effects may provide an explanation for why plants seem to be more phenotypically plastic than animals. Relatively fluid cytosine methylation signatures provide a large source of environmentally-alterable additive genetic variance that has previously largely been disregarded. Meiotic genomes primarily only have methylation on their cytosine nucleotides that are in CpG dinucleotides (cytosine-phosphate-guanine) or, much more rarely, in CpNpG trinucleotides, where N can be any nucleotide (Ramsahoye et al., 2000). Animals generally having less than 5% of their CpG dinucleotides methylated, whereas plants have 10-30% of their CpG dinucleotides methylated (Shapiro, 1976; Finnegan & Kovac, 2000). A test of the assertion that cytosine methylation fosters phenotypic plasticity would be to see whether the highest incidences of phenotypic plasticity occur in those taxa with high cytosine methylation levels and that the highest incidences of canalization occur in those taxa with little or no cytosine methylation. We would have to examine a broad array of plant and animal taxa, realizing that not all environmentally-alterable epigenetic signals are based on cytosine methylation (e.g. histone acetylation; Prahald et al., 2003).

For those environmentally-alterable additive genetic effects that are due to epigenetic signals, Waddington (1957: ix) showed great foresight in saying that, "Neo–Darwinism involves a breach between organism and nature as complete as the Cartesian

dualism of mind and matter; an epigenetic consideration of evolution goes some way towards healing it."

There are both theoretical and practical implications of this work. Evolutionary theorists are always looking for mechanisms that increase additive genetic variance. Here that is done by converting environmental variance to additive genetic variance. Some pragmatic implications of this work are discussed in Appendix A. For example, if pollutants adversely affect organisms by heritably altering their epigenetic signals, then environmental remediation will not remove the problem. Once pollutants have adversely altered an individual's epigenetic signals, this harm will be transmitted to future generations even if they are not exposed to the pollutants. It should become vitally important to estimate this environmentally-alterable component, V_{EA}, of additive genetic variance. In fact, a primary goal of this paper is to encourage empirical quantitative geneticists to estimate V_{EA}. Assuming that the interaction variances are zero (i.e. V_{GxE}, $V_{GxE\text{-parent}}$, and epistatic variance are zero), if V_{EA} is zero, then we can decompose phenotypic variance into components due to nature (V_A) and nurture (V_C+V_E), respectively. If, however, V_{EA} is a substantial portion of V_P, especially compared with V_A, then it will be impossible to disentangle the effects of nature and nurture.

CHAPTER 5

DISCUSSION

I have developed several mathematical and verbal aspects of evolutionary epigenetic theory. I affirmed that standard Fisherian quantitative genetics are adequate to describe epigenetic phenomena. Therefore, epigenetic effects can be modeled as epistatic effects or as environmentally-alterable additive genetic effects. Likewise, although I have not shown this here, the epigenetic signals of genomic imprinting can be modeled in quantitative genetic models as Mendelian effects by focusing on the differences between female and male epigenetic signals (Gorelick & Bertram, submitted), using techniques developed in Lande (1979).

By asserting that epigenetic effects are synonymous with epistatic genetic effects, I was able to model signals that molecular biologists call constitutive epigenetic. Thus, even without the fluidity that is their trademark, epigenetic effects contribute to additive genetic variance. By asserting that other epigenetic effects are essentially fluid genetic effects – i.e. are less heritable than nucleotides – I was able to model signals that molecular biologists call facultative epigenetic. I showed that fluidity of meiotically-heritable epigenetic signals that is due to environmental perturbations can be modeled as environmentally-alterable additive genetic effects, which is a generalization of indirect additive genetic effects. This provides a second mechanism by which epigenetic effects contribute to additive genetic variance. And, it does so in a way that renders evolutionary trajectories entirely contingent on the time history of the environment. Both constitutive

and facultative epigenetic effects contribute to additive genetic variance, although this has never been previously realized except in the context of developmental modules. What this work will allow us to do is quantify – both empirically and theoretically – the relative contributions the effects of these epigenetic effects on additive genetic variance (something we currently do not know).

The evolutionary epigenetic theory in chapters 1, 2, and 4 also sheds light on what is meant by a gene. Because epigenetic signals such as cytosine methylation and chromatin formation can contribute to additive genetic variance, we can no longer readily exclude them the definition of a gene. Even assuming that there are no epistatic effects other than the interaction of single epigenetic loci with their underlying single genetic loci, epigenetic signals should either be considered genes or parts of genes. This assertion is consistent with Benzer's cis-trans definition of genes (Benzer, 1959), but diverges from more contemporary definitions which state that genes are lengths of nucleotides (e.g. Kitchner, 1992; Sarkar, 1998; Portin, 2002). My work supports a return to earlier notions of the meaning of genes simply as units of heredity (Johannsen, 1909 as cited in Maienschein, 1992) or as something that contributes to additive genetic variance (Fisher, 1930; Lush, 1937). In other words, a gene is merely something that is heritable or codes for a heritable trait. This is a quintessential black box definition of genes, but one that seems to be required if we are to incorporate epigenetic effects into a neo-Darwinian framework. Like Fisher, we need a black box approach to the gene and the genotype-to-phenotype map in order to develop tractable mathematical models of evolution. Only after that should we develop more reductionist models that explicitly

incorporate the molecular biology that underlies genetic/epigenetic effects and the genotype-to-phenotype map – but subsume the black box approach – which should lead to some interesting and unexpected theoretical results.

I applied the work on environmentally-alterable additive genetic effects to fluid epigenetic signals to postulate one possible molecular mechanism underlying G x E interactions and phenotypic plasticity. Although at first these seem like abstruse mathematical arguments, I showed that this new perspective can have huge implications for public health and environmental policy. In particular, it begs for prevention rather than cure of cancers and other diseases (see Parenti, 2003 for a more political, but equally valid argument regarding asthma). Environmentally-alterable additive genetic effects that are caused by fluid epigenetic signals may also provide the reason for why plants are so much more phenotypically plastic than animals – plants have much higher concentrations of cytosine methylation than animals, which may in part be due to their higher incidence of polyploidy.

The usefulness of the theory developed in chapters 2 and 4 and Appendix A will be dependent upon an open empirical question: How heritable are epigenetic signals? I have described and cited copious qualitative evidence for the epigenetic signals of cytosine methylation and chromatin formation (including genomic imprinting) being highly heritable, albeit less so than nucleotides. These epigenetic signals are known to be transmitted to several successive subsequent generations, although here empirical work is largely lacking. Consistent chromosomal banding patterns within a lineage over both ontogeny and phylogeny provide the best existing evidence that these epigenetic signals

are heritable over many successive generations. If further empirical work corroborates and quantifies this heritability, then the evolutionary epigenetic theory developed herein could be quite important. The applicability of this theory would grow even more if other forms of epigenetic signals such as RNA editing mechanisms, rRNA, MTOCs, etc, are found to also be heritable over many successive generations.

Appendix B, C, and D show theoretical implications for evolutionary epigenetic theory in the origins and evolution of sex. I have spent much of the past two years deciphering the roles evolutionary epigenetics have played in the origins of meiosis, sex determination, dioecy, sex chromosomes, and sexual dimorphism. Although I am still putting the pieces together and gathering more evidence (e.g. dioecy really does seem to be triggered by polyploidy; Miller & Venable, 2000), hard work still remains, especially on processes that occurred further back in geologic time, such as any role that cytosine methylation may have played in the origins of meiosis.

To most people, my title "evolutionary epigenetic theory" invokes notions of developmental biology, especially the active research areas known as evo-devo (e.g. Cowley & Atchley, 1992; Hall, 1992). Classic evo-devo ideas lurk in this dissertation, although sometimes in cryptic fashion. Appendix A and C contain discussions of how epigenetic signals may affect differentiation and de-differentiation of tissues in animals and plants, respectively. Appendix D discusses how ancient asexual animal lineages probably go through most of the substantive life history stages of related sexual animals. In particular, ancient asexual animals probably undergo meiosis in the form of automixis, and can thereby periodically reset their epigenetic signals. Appendix B presents the most

compelling evo-devo tale insofar as cytosine methylation provides the source for canalization of dioecy and sex chromosomes in animals. Cytosine methylation triggers this canalization by driving Muller's ratchet, which (like most ratchets) is effectively irreversible.

Gene duplications induce epigenetic changes in cytosine methylation and chromatin formation (Ohno, 1970; Holland, Garcia-Fernandez, Williams & Sidow, 1994; Cooke, Nowak, Boerlijst & Maynard Smith, 1997; Holland, 1998; Force, Lynch, Pickett, Amores, Yan & Postlethwait, 1999). I showed how these epigenetic effects could have been the proximate cause for the origins of sex determination, dioecy, and sex chromosomes. These gene-duplication-induced epigenetic changes canalized dioecy and sex chromosomes. I therefore assert that evolution of dioecy can be construed as a peculiar form of neoteny in which half of each lineage matures before producing eggs and the other half matures before producing sperm (although, admittedly, I have never before seen this re-conceptualization of dioecy).

On curious thing about the model in Appendix B is that short-circuiting Muller's ratchet means that canalization should not occur. This, in fact, appears to be the case in plants, which experience so much selection in their haploid phases that they are immune from Muller's ratchet. Plants thus are less canalized and more phenotypically plastic than animals with respect to sex determination. Compared with animals, elevated levels of phenotypic plasticity also seem to be prevalent in other plant traits. Could this be due to greater concentrations of cytosine methylation in plants, which could then be environmentally-altered? Could the extra methylation itself be a result of greater

incidences of polyploidy? My guess is 'yes' to both questions (see the next paragraph for some details), although an argument can be made that plants are more tolerant of polyploidy because they are more phenotypically plastic. I have argued elsewhere (Gorelick, 2001) that the way to disentangle these effects is to compare seed plant lineages that incur polyploidization with those that do not, but have not yet conducted such an analysis.

I am beginning to develop some general theory that polyploidy can induce paedomorphism, including neoteny and progenesis. The possibility exists that polyploidy can trigger new epigenetic signatures that down-regulate those genes that regulate latter stages of development (not just sexual function). If the nascent epigenetic signatures can be removed via environmental perturbations, then this results in progenesis. If the nascent epigenetic signatures are more permanent, then this results in neoteny. I am beginning to assemble evidence for this idea as well as develop the theory more explicitly. These ideas seem to be borne out by salamanders (Roth, Nishikawa, Naujoksmanteuffel, Schmidt & Wake, 1993; Roth, Nishikawa & Wake, 1997) and seed plants (Liu & Wendel, 2003; Gorelick, submitted-a). Earlier, I hypothesized that several diminutive South American geophytic cacti seem to be neotenous and that this was triggered by polyploidy (Gorelick, 2004). Patrick Griffith, Ralph Peters, and I are planning on testing this hypothesis using flow cytometry on the genera *Pediocactus* and *Sclerocactus* in northwest Arizona (and neighboring portions of southern Nevada, southern Utah, and northwestern New Mexico). I have classified taxa as normal (x=0), progenetic (x=1), and neotenic (x=2). Using phylogenetic comparative methods (Patrick

Griffith is currently constructing a phylogeny based on molecular and morphological characters), I predict that levels of paedomorphism will be correlated with levels of polyploidy.

In chapter 2, I showed how the effects of epigenetic signals sensu Holliday (e.g. cytosine methylation, chromatin formation) were mathematically equivalent physiological epistasis. Ideally, what I would like to do is show that, within a Fisherian framework, all epigenetic signals (including the ones envisioned by Waddington and people currently working on evo-devo) are equivalent to physiological epistasis. I am also trying to show that indirect genetic effects can also be subsumed in models of physiological epistasis. This would provide the noble goal of incorporating all forms of epistatic and epigenetic effects under a single umbrella of a formal Fisherian model.

There is enormous potential for advancing evolutionary theory by meshing epigenetics with population and quantitative genetics. If epigenetics provide a sizeable source for additive genetic variance, biologists will gain insight into the fuel that fires evolution. This could eliminate the evolutionary cul-de-sac predicted by Fisher's fundamental theorem of natural selection, without having to invoke unrealistic levels of mutation or other *deus ex machina*. Other consequences of this evolutionary epigenetic theory can be as mundane as explaining the evolution of sex chromosomes and the maintenance of obligately asexual animal lineages for tens or millions of years. The consequences of this evolutionary epigenetic theory can also be of practical importance to many people in society: from being able to exclusively grow female marijuana plants to preventing cancer. The questions, few answers, and research agenda that I outlined

above only begin to scratch the surface of applying existing genetic techniques to epigenetics. Such an epigenetic frontier establishes a theme that could be productively pursued for many years and could help unify evolutionary theory.

REFERENCES

Abbott EA. 1899. *Flatland*. Little Brown, Boston.

Andrews GE. 1976. *The theory of partitions*. Addison-Wesley, Reading.

Ardlie KG, Kruglyak L, and Seielstad M. 2002. Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* **3:** 299-309.

Avery OT, MacLeod CM, and McCarty M. 1944. Studies on the chemical nature of the substance inducing transformation in pneumococcal types. *Journal of Experimental Medicine* **79:** 137-158.

Beck S, Olek A, and Walter J. 1999. From genomics to epigenomics: a loftier view of life. *Nature Biotechnology* **17:** 1144.

Bennett JH. 1954. On the theory of random mating. *Annals of Eugenics* **18:** 311-317.

Benzer S. 1959. On the topology of the genetic fine structure. *Proceedings of the National Academy of Sciences of the United States of America* **45:** 1607-1620.

Brachet J. 1940. La détection histochimique des acides pentosenucléiques. *Comptes Rendus des Séances de la Société Belge de Biologie* **133:** 88-90.

Bulmer MG. 1972. The genetic variability of polygenic characters under optimizing selection, mutation and drift. *Genet. Res.* **19:** 17-25.

Bulmer MG. 1980. *The mathematical theory of quantitative genetics*. Clarendon Press, Oxford.

Califia P. 1988. *Macho Sluts*. Alyson Publications, Boston.

Caspersson T, and Schultz J. 1940. Nucleic acid metabolism of the chromosomes in relation to gene reproduction. *Proceedings of the National Academy of Sciences of the United States of America* **26:** 507-515.

Cech TR, and Bass BL. 1986. Biological catalysis by RNA. *Annual Review of Biochemistry* **55:** 599-629.

Champagne FA, and Meaney MJ. 2001. Like mother, like daughter: evidence for non-genomic transmission of parental behavior and stress responsivity. *Progress in Brain Research* **133:** 287-302.

Cheverud JM, and Routman EJ. 1995. Epistasis and its contribution to genetic variance components. *Genetics* **139:** 1455-1461.

Cooke J, Nowak MA, Boerlijst M, and Maynard Smith J. 1997. Evolutionary origins and maintenance of redundant gene expression during metazoan development. *Trends in Genetics* **13:** 360-364.

Cowley DE, and Atchley WR. 1992. Quantitative genetic models for development, epigenetic selection, and phenotypic evolution. *Evolution* **46:** 495-518.

Crick F. 1970. Central dogma of molecular biology. *Nature* **227:** 561-563.

Croteau S, Andrade MF, Huang F, Greenwood CMT, Morgan K, and Naumova AK. 2002. Inheritance patterns of maternal alleles in imprinted regions of the mouse genome at different stages of development. *Mammalian Genome* **13:** 24-29.

Crouse HV. 1960. The controlling element in sex chromosome behavior in *Sciara*. *Genetics* **45:** 1429-1443.

Darwin CR. 1872. *The origin of species by means of natural selection (6th edition)*. John Murray, London.

Darwin CR. 1875. *The variation of animals and plants under domestication (2nd edition)*. John Murray, London.

Dausset J, Legrand L, Lepage V, Contú L, Marcelli-Barge A, Wildloecher I, Benajam A, Meo T, and Degos L. 1978. A haplotype study of HLA complex with special reference to HLA-DR series and to Bf, C2 and glyoxalase I polymorphisms. *Tissue Antigens* **12:** 297-307.

Dobzhansky T. 1948. Genetics of natural populations. XVIII. Experiments on chromosomes of *Drosophila pseudoobscura* from different geographical regions. *Genetics* **33:** 797-808.

Dorazi R, Chesnel A, and Dournan C. 1995. Opposite sex determination of gonads in two *Pleurodeles* species may be due to temperature-dependent inactivation of sex chromosomes. *Journal of Heredity* **86:** 28-31.

Ehrlich M. 2000. DNA methylation: normal development, inherited diseases, and cancer. *Journal of Clinical Ligand Assay* **23:** 144-146.

Ekwall K, Olsson T, Turner BM, Cranston G, and Allshire RC. 1997. Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. *Cell* **91:** 1021-1032.

Falconer DS, and Mackay TFC. 1996. *Introduction to quantitative genetics (4th edition)*. Prentice Hall, Harlow.

Ficq A, and Pavan C. 1957. Autoradiography of polytene chromosomes of *Rhynchosciara angelae* at different stages of larval development. *Nature* **180:** 983-984.

Finnegan EJ, and Kovac KA. 2000. Plant DNA methyltransferases. *Plant Molecular Biology* **43:** 189-201.

Fisher RA. 1930. *The genetical theory of natural selection*. Clarendon Press, Oxford.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, and Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151:** 1531-1545.

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, and Paul CL. 1992. A genomic sequencing protocol which yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America* **89:** 1827-1831.

Geiringer H. 1944. On the probability theory of linkage in Mendelian heredity. *Annals of Mathematical Statistics* **15:** 25-57.

Gilbert W. 1986. The RNA world. *Nature* **319:** 618-618.

Goodnight CJ. 1988. Epistasis and the effect of founder events on the additive genetic variance. *Evolution* **42:** 441-454.

Goodnight CJ. 1995. Epistasis and the increase in additive genetic variance: implications for phase 1 of Wright's shifting balance process. *Evolution* **49:** 502-511.

Goodnight CJ. 2000. Modeling gene interaction in structured populations. In: Wolf JB, Brodie ED and Wade MJ, eds. *Epistasis and the evolutionary process*. Oxford: Oxford University Press. 129-145.

Gorelick R. 2001. Did insect pollination cause increased seed plant diversity? *Biological Journal of the Linnean Society* **74:** 407-427.

Gorelick R. 2003. Evolution of dioecy and sex chromosomes via methylation driving Muller's ratchet. *Biological Journal of the Linnean Society* **80:** 353-368.

Gorelick R. 2004. Resolving the phylogenetic placement of *Blossfeldia liliputana* (Cactaceae): reticulate evolution, chloroplast inheritance, and graft-chimeras. *Bradleya* **22:** 9-14.

Gorelick R. submitted-a. Is lack of cycad diversity due to lack of polyploidy?

Gorelick R. submitted-b. Why are there two global de novo methylation events each generation? Coevolution of epigenetic resets and meiotic reduction divisions.

Gorelick R, and Bertram SM. submitted. Evolution of sexually dimorphic traits and maternal inheritance.

Gorelick R, and Osborne R. 2002. Inducing sex change and organogenesis from tissue culture in the endangered African cycad *Encephalartos woodii* (Cycadales, Zamiaceae). *South African Journal of Science* **98:** 114-117.

Gowher H, Leismann O, and Jeltsch A. 2000. DNA of *Drosophila melanogaster* contains 5-methylcytosine. *EMBO Journal* **19:** 6918-6923.

Greene WH. 1997. *Econometric analysis (3rd edition)*. Prentice-Hall, Upper Saddle River.

Griffith MP. 2004. The origins of an important cactus crop, *Opuntia ficus-indica* (Cactaceae): new molecular evidence. *American Journal of Botany* **in press**.

Hall BK. 1992. *Evolutiuonary developmental biology*. Chapman & Hall, London.

Hastings A. 1984. Linkage disequilibrium, selection and recombination at three loci. *Genetics* **106:** 153-164.

Hedrick PW. 2000. *Genetics of populations (2nd edition)*. Jones and Bartlett, Sudbury.

Hershey AD, and Chase M. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology* **36:** 39-56.

Hogben L. 1933. *Nature and nurture*. W. W. Norton, New York.

Holland PWH. 1998. Major transitions in animal evolution: a developmental genetic perspective. *American Zoologist* **38:** 829-842.

Holland PWH, Garcia-Fernandez J, Williams NA, and Sidow A. 1994. Gene duplications and the origins of vertebrate development. *Development* **1994 Supplement:** 125-133.

Holliday R. 1987. The inheritance of epigenetic defects. *Science* **238:** 163-170.

Holliday R. 1994. Epigenetics: an overview. *Developmental Genetics* **15:** 453-457.

Holliday R, and Pugh JE. 1975. DNA modification mechanisms and gene activity during development. *Science* **187:** 226-232.

Iyengar VK, Reeve HK, and Eisner T. 2002. Paternal inheritance of a female moth's mating preference. *Nature* **419:** 830-832.

Jenuwein T. 2001. Re-SET-ting heterochromatin by histone methyltransferases. *Trends in Cell Biology* **11:** 266-273.

Johannsen W. 1909. *Elemente der exakten erblichkeitslehre*. Gustav Fischer, Jena.

Jollos V. 1934. Inherited changes produced by heat treatment in *Drosophila*. *Genetica* **16:** 476-494.

Judson HF. 1979. *The eighth day of creation*. Simon & Schuster, New York.

Keller EF. 1983. *A feeling for the organism: the life and work of Barbara McClintock*. W. H. Freeman, New York.

Kirkpatrick M, and Lande R. 1989. The evolution of maternal characters. *Evolution* **43:** 485-503.

Kitchner P. 1992. Gene: current usage. In: Keller EF and Lloyd EA, eds. *Keywords in evolutionary biology*. Cambridge: Harvard University Press. 128-131.

Lakoff G, and Núñez RE. 2000. *Where mathematics comes from*. Basic Books, New York.

Lande R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* **33:** 402-416.

Lande R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution* **34:** 292-305.

Laubichler MD. 1997. Identifying units of selection: conceptual and methodological issues. New Haven: Yale University.

Lewontin RC. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49:** 49-67.

Lewontin RC. 1974. *The genetic basis of evolutionary change*. Columbia University Press, New York.

Lewontin RC. 1991. *Biology as ideology*. Harper Collins, New York.

Lewontin RC. 1992. Genotype and phenotype. In: Keller EF and Lloyd EA, eds. *Keywords in evolutionary biology*. Cambridge: Harvard University Press. 137-144.

Liu B, and Wendel JF. 2003. Epigenetic phenomena and the evolution of plant allopolyploids. *Molecular Phylogenetics and Evolution* **29:** 365-379.

Lush JL. 1937. *Animal breeding plans*. Iowa State College Press, Ames.

Lush JL. 1940. Inter-sire correlations or regressions of offspring on dam as a method of estimating heritability of characters. *American Society of Animal Production, Proceedings* **33:** 293-301.

Lyko F, Ramsahoye BH, and Jaenisch R. 2000. DNA methylation in *Drosophila melanogaster*. *Nature* **408:** 538-540.

Lynch M. 1991. The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution* **45:** 622-629.

Lynch M, and Walsh B. 1998. *Genetics and analysis of quantitative traits*. Sinauer Associates, Sunderland.

Maienschein J. 1992. Gene: historical perspectives. In: Keller EF and Lloyd EA, eds. *Keywords in evolutionary biology*. Cambridge: Harvard University Press. 122-127.

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

Mayer W, Smith A, Fundele R, and Haaf T. 2000. Spatial separation of parental genomes in preimplantation mouse embryos. *Journal of Cell Biology* **148:** 629-634.

McClintock B. 1950. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America* **36:** 344-355.

Meaney MJ. 2001. Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience* **24:** 1161-1192.

Miller JS, and Venable DL. 2000. Polyploidy and the evolution of gender dimorphism in plants. *Science* **289:** 2335-2338.

Miller OJ, Schnedl W, Allen J, and Erlanger BF. 1974. 5-methylcytosine localised in mammalian constitutive heterochromatin. *Nature* **251:** 636-637.

Miniou P, Jeanpierre M, Blanquet V, Sibella V, Bonneau D, Herbelin C, Fischer A, Niveleau A, and Viegas-Péquignot E. 1994. Abnormal methylation pattern in constitutive and facultative (X-inactive chromosome) heterochromatin of ICF patients. *Human Molecular Genetics* **3:** 2093-2102.

Moore AJ, Brodie ED, and Wolf JB. 1997. Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. *Evolution* **51:** 1352-1362.

Ohno S. 1970. *Evolution by gene duplication*. Springer-Verlag, Berlin.

Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R, Dean W, Reik W, and Walter J. 2000. Active demethylation of the paternal genome in the mouse zygote. *Current Biology* **10:** 475-478.

Palopoli MF, and Wu CI. 1996. Rapid evolution of a coadapted gene complex: evidence from the segregation distorter (SD) system of meiotic drive in *Drosophila melanogaster*. *Genetics* **143:** 1675-1688.

Parenti M. 2003. Media moments. In: Phillips P and Project-Censored, eds. *Censored 2004*. New York: Seven Stories Press. 125-128.

Pollock G. 2001. *Assays in American constitutional process*. Unpublished treatise.

Poole A, Penny D, and Sjöberg BM. 2000. Methyl-RNA: An evolutionary bridge between RNA and DNA? *Chemistry & Biology* **7:** R207-R216.

Poole A, Penny D, and Sjöberg B-M. 2001. Confounded cytosine! Tinkering and the evolution of DNA. *Nature Reviews Molecular Cell Biology* **2:** 147-151.

Portin P. 2002. Historical development of the concept of the gene. *Journal of Medicine and Philosophy* **27:** 257-286.

Prahald V, Pilgrim D, and Goodwin EB. 2003. Roles of mating and environment in *C. elegans* sex determination. *Science* **302:** 1046-1049.

Rakyan V, and Whitelaw E. 2003. Transgenerational epigenetic inheritance. *Current Biology* **13:** R6.

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, and Jaenisch R. 2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated

by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 5237-5242.

Rawson PD, and Burton RS. 2002. Functional coadaptation between cytochrome c and cytochrome c oxidase within allopatric populations of a marine copepod. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 12955-12958.

Riggs AD. 1975. X inactivation, differentiation, and DNA methylation. *Cytogenetics and Cell Genetics* **14:** 9-11.

Riggs AD, Martienssen R, and Russo VEA eds. 1996a. *Epigenetic mechanisms in gene regulation*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Riggs AD, Martienssen R, and Russo VEA. 1996b. Introduction. In: Russo VEA, Martienssen R and Riggs AD, eds. *Epigenetic mechanisms in gene regulation*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press. 1-4.

Roff DA. 1986. The genetic basis of wing dimorphism in the sand cricket, Gryllus firmus and its relevance to the evolution of wing dimorphisms in insects. *Heredity* **57:** 221-231.

Roth G, Nishikawa KC, Naujoksmanteuffel C, Schmidt A, and Wake DB. 1993. Paedomorphosis and simplification in the nervous system of salamanders. *Brain Behavior and Evolution* **42:** 137-170.

Roth G, Nishikawa KC, and Wake DB. 1997. Genome size, secondary simplification, and the evolution of the brain in salamanders. *Brain Behavior and Evolution* **50:** 50-59.

Sarkar S. 1998. *Genetics and reductionism*. Cambridge University Press, Cambridge.

Scheiner SM, and Goodnight CJ. 1984. The comparison of phenotypic plasticity and genetic variation in populations of the grass *Danthonia spicata*. *Evolution* **38:** 845-855.

Schlichting CD, and Pigliucci M. 1998. *Phenotypic evolution: a reaction norm perspective*. Sinauer, Sunderland.

Schluter D. 1998. Ecological causes of speciation. In: Howard DJ and Berlocher SH, eds. *Endless forms: species and speciation*. Oxford: Oxford University Press. 114-129.

Serfling E, Jasin M, and Schaffner W. 1985. Enhancers and eukaryotic gene transcription. *Trends in Genetics* **1:** 224-230.

Shapiro HS. 1976. Distribution of purines and pyrimidines in deoxyribonucleic acids. In: Fastman GD, ed. *Handbook of biochemistry and molecular biology. Volume II. Nucleic acids*. 3rd ed. Cleveland: CRC Press. 241-281.

Solari AJ. 1994. *Sex chromosomes and sex determination in vertebrates*. CRC Press, Boca Raton.

Sturtevant AH. 1925. The effects of unequal cross-over at the *Bar*-locus in *Drosophila*. *Genetics* **10:** 117-147.

Tatra GS, Miranda J, Chinnappa CC, and Reid DM. 2000. Effect of light quality and 5-azacytidine on genomic methylation and stem elongation in two ecotypes of *Stellaria longipes*. *Physiologia Plantarum* **109:** 313-321.

Thomson G, and Baur MP. 1984. Third order linkage disequilibrium. *Tissue Antigens* **24:** 250-255.

Traut W, and Marec F. 1996. Sex chromatin in Lepidoptera. *Quarterly Review of Biology* **71:** 239-256.

Tuck-Muller CM, Narayan A, Tsien F, Smeets DFCM, Sawyer J, Fiala ES, Sohn OS, and Ehrlich M. 2000. DNA hypomethylation and unusual chromosome instability in cell lines from ICF syndrome patients. *Cytogenetics and Cell Genetics* **89:** 121-128.

Waddington CH. 1940. *Organizers and genes*. Cambridge University Press, Cambridge.

Waddington CH. 1957. *The strategy of the genes: a discussion of some aspects of theoretical biology*. George Allen & Unwin, London.

Wade MJ, and Goodnight CJ. 1998. Perspective: The theories of Fisher and Wright in the context of metapopulations: When nature does many small experiments. *Evolution* **52:** 1537-1553.

Wade MJ, Winther RG, Agrawal AF, and Goodnight CJ. 2001. Alternative definitions of epistasis: dependence and interaction. *Trends in Ecology & Evolution* **16:** 498-504.

Wagner GP, and Laubichler MD. 2000. Character identification in evolutionary biology: the role of the organism. *Theory in Biosciences* **119:** 20-40.

Wagner GP, Laubichler MD, and Bagheri-Chaichian H. 1998. Genetic measurement theory of epistatic effects. *Genetica* **102/103:** 569-580.

Waugh O'Neill RJ, Eldridge MDB, and Graves JAM. 2001. Chromosome heterozygosity and de novo chromosome rearrangements in mammalian interspecies hybrids. *Mammalian Genome* **12:** 256-259.

Waugh O'Neill RJ, O'Neill MJ, and Graves JAM. 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* **393:** 68-72.

Weaver ICG, Szyf M, and Meaney MJ. 2002. From maternal care to gene expression: DNA methylation and the maternal programming of stress responses. *Endocrine Research* **28:** 699.

Williams GC. 1988. Retrospect on sex and kindred topics. In: Michod RE and Levin BR, eds. *The evolution of sex: an examination of current ideas*. Sunderland: Sinauer Associates. 287-298.

Wolf JB. 2003. Genetic architecture and evolutionary constraint when the environment contains genes. *Proceedings of the National Academy of Sciences of the United States of America* **100:** 4655-4660.

Wolf JB, and Brodie ED. 1998. The coadaptation of parental and offspring characters. *Evolution* **52:** 299-308.

Wolf JB, Brodie ED, Cheverud JM, Moore AJ, and Wade MJ. 1998. Evolutionary consequences of indirect genetic effects. *Trends in Ecology & Evolution* **13:** 64-69.

Wolf JB, Brodie ED, and Wade MJ eds. 2000. *Epistasis and the evolutionary process*. Oxford: Oxford University Press.

Wolf JB, Frankino WA, Agrawal AF, Brodie ED, and Moore AJ. 2001. Developmental interactions and the constituents of quantitative variation. *Evolution* **55:** 232-245.

Wolff GL, Kodell RL, Moore SR, and Cooney CA. 1998. Maternal epigenetics and methyl supplements affect *agouti* gene expression in A^{vy}/a mice. *FASEB Journal* **12:** 949-957.

Wolffe AP, and Matzke MA. 1999. Epigenetics: regulation through repression. *Science* **286:** 481-486.

Wright S. 1931. Evolution in Mendelian populations. *Genetics* **16:** 97-159.

APPENDIX A

NEO-LAMARCKIAN MEDICINE

Medical Hypotheses (2004) **62**, 299–303

ELSEVIER

medical hypotheses

http://intl.elsevierhealth.com/journals/mehy

Neo-Lamarckian medicine

Root Gorelick*

School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501, USA

Received 18 August 2003; accepted 4 November 2003

Summary Darwinian medicine is the treatment of disease based on evolution. The underlying assumption of Darwinian medicine is that traits are coded by genes, which are often assumed to be sequences of DNA nucleotides. The quantitative genetic ramification of this perspective is that traits, including disease susceptibility, are either caused by genes or by the environment, with genotype-by-environment interactions usually considered statistical artefacts. I emphasize also examining those epigenetic signals that can be altered by environmental perturbations and then transmitted to subsequent generations. Although seldom studied, environmentally-alterable meiotically-heritable epigenetic signals exist and provide a mechanism underlying genotype-by-environment interactions. Environment of a parent can affect its descendants by heritably altering epigenetic signals. Neo-Lamarckian medicine is the application of these evolutionary epigenetic notions to diseases and could have enormous public health and environmental policy implications. If industrial contaminants adversely affect organisms by meiotically-heritably altering their epigenetic signals, then cleaning up these contaminants will not remedy the problem. Once contaminants have adversely altered an individual's epigenetic signals, this harm will be transmitted to future generations even if they are not exposed to the contaminant. Exposure to environmental shocks such as free radicals or other carcinogens can alter cytosine methylation patterns on regulatory genes. This can cause cancer by up-regulating genes for cell division or by down-regulating tumour suppressor genes. Environmentally-alterable meiotically-heritable epigenetic signals could also underlie other diseases, such as diabetes, Prader–Willi syndrome, and many complex diseases. If environmentally-altered meiotically-heritable epigenetic effects are widespread – which is an important open empirical question – they have the potential to alter paradigmatic views of evolutionary medicine and the putative dichotomy of nature versus nurture. Neo-Lamarckian medicine would thereby shift emphasis from cure to prevention of diseases.

Introduction

Darwinian medicine has captured the imagination of medical researchers and evolutionary biologists [1,2]. It is defined as the use of "an evolutionary perspective to understand why the body is not better designed and why, therefore, diseases exist at all" [3, p. 358]. The term 'Darwinian' is used to distinguish neo-Darwinian views of evolution from other, often older, views of evolution. In particular, Darwinian evolution is often contrasted with Lamarckian notions of evolution of acquired characters, which are clearly wrong. However, there are portions of cells other than DNA that are faithfully transmitted from one generation to the next, and some of these can be heritably altered by the environment [4]. These signals provide a mode by which the environment in one generation can affect the evolutionary trajectory of subsequent generations. Jablonka and Lamb have discussed many such evolutionary applications, but none in medicine [4–6]. I fill in this gap by introducing such epigenetic effects into evolutionary medicine, calling them neo-Lamarckian medicine.

Neo-Lamarckian medicine is driven by epigenetic signals that are inherited across generations, but are more fluid than DNA nucleotides. Unfortunately, no term exists for such signals, hence I begin by naming these 'meiotically-heritable

* Tel.: +1-480-966-1901; fax: +1-480-965-2519.
E-mail address: cycad@asu.edu (R. Gorelick).

0306-9877/$ - see front matter
doi:10.1016/S0306-9877(03)00329-3

epigenetic' signals to distinguish them from epigenetic that are only transmitted via mitosis.

What are meiotically-heritable epigenetic signals?

The term epigenetic was originally used to describe all factors controlling gene expression and cell differentiation. Waddington [7] used the term epigenetic, amalgamating epigenesis and genetics, where epigenesis refers to the gradual and progressive development of new structures [8]. It was not until Waddington was writing his seminal book that biologists realized that DNA is contained in chromosomes ([9] citing [10]). Genetic phenomena, like epigenetic ones, were originally considered to be of unknown molecular cause. Only genetic and epigenetic effects were known. Fisher defined additive genetic variance as the proportion of phenotypic effects that get transmitted from one generation to the next [11]. Note that this is a population level definition of genetic effects because it requires one to examine average effects and variances. In contrast, Waddington used the term epigenetic to refer to any phenomena that affected development *within an individual* and, therefore implicitly, *within a generation*.

The meaning of the term genetic changed with the introduction of the central dogma and sequence hypothesis [12]. They provided a one-way path from DNA nucleotide sequences, to messenger RNA (transcription), to protein synthesis (translation). Because of this paradigm, molecular biologists effectively re-defined the term genetic to mean DNA nucleotide sequences; a definition that many quantitative geneticists would admittedly find repugnant.

Use of the term epigenetics also changed in the 1970s as molecular biologists began proposing mechanisms, especially cytosine methylation, by which development and gene regulation could be controlled [13,14]. Many of these molecular signatures gradually change irreversibly as cells and tissues differentiate and age. Therefore, whichever molecular mechanisms control development must be reset each generation. "This suggests a supplementary definition of epigenetics to include transmission from one generation to the next, other than the DNA sequence itself" [15, p. 454]. Holliday thereby switched the focus of epigenetics from strictly intra-generational to both intra- and inter-generational (also see [16]). Therefore, I use the terms 'mitotically-heritable' and 'meiotically-heritable' to distinguish the two different contemporary meanings of the term epigenetic. Meiotically-heritable epigenetic signals include cytosine methylation, chromatin structure, and histone acetylation and undoubtedly also include other molecular signals, such as those mediating RNA editing.

Do environmentally-altered meiotically-heritable epigenetic signals exist?

There are several impediments to providing a definitive answer as to whether there exist environmentally-alterable meiotically-heritable e-pigenetic signals. First, most work on heritability of epigenetic signals has focused on faithful transmission of epigenetic signals through mitosis, not meiosis. Second, although it is clear that some epigenetic signals — such as the cytosine methylation signatures of genomic imprinting and heterochromatin — are meiotically-heritable, nobody has quantified their degree of meiotic heritability (i.e. narrow-sense heritability). Third, methods for estimating these environmentally-altered epigenetic components of heritability have not been developed. To remedy this third difficulty, I provide a sketch for an estimation methodology below (details contained in [17]). Neo-Lamarckian medicine can only exist if environmentally-altered meiotically heritable epigenetic signals exist.

At least six labs have independently reported environmentally-altered meiotically-heritable epigenetic signals, two of which work with cytosine methylation in mice. At some loci, amount of gene product is proportional to amount of methylation on the promoter [18]. At one mouse locus, methylation levels and resulting phenotypic effects were both meiotically-heritable and could be altered by feeding maternal parents methyl-rich diets [19]. Similar results were found at another mouse locus [20], but here epigenetic signals were meiotically-heritable and could be altered through either maternal or paternal parents. Chemical alteration of cytosine methylation patterns is also known to be meiotically-heritable in plants [21,22]. Environmentally-alterable meiotically-heritable epigenetic signals are not limited to cytosine methylation. They has also been found in fission yeast [23] and *Drosophila* [24], which are two of the few lineages that has independently lost all or most of their cytosine methylation [25]. We cannot know how common or important these phenomena are until researchers become more cognizant of and systematically search for environmentally-altered meiotically-heritable epigenetic signals [6].

Quantitative genetics of environmentally-altered meiotically-heritable epigenetics

It is surprising that meiotically-heritable phenotypic traits are usually exclusively attributed to genetic and environmental causes, and not also to epigenetic causes. There has been a constant epigenetic thread running through evolutionary biology since the discovery of position effect variegation in the 1920s [26], transposons in the 1940s [27], genomic imprinting in the 1960s [28], and the regulatory roles of cytosine methylation in the 1970s [14]. Each of these phenomena should have injected an epigenetic element into quantitative genetic paradigms, but somehow failed to do so. Thus far, work has been exclusively confined to using quantitative genetics to estimate developmental epigenetics [29,30].

Quantitative genetics considers genetic effects to be any portion of variation that cannot be explained by environmental variation. Parent-offspring models are linear regressions,where offspring phenotype is the dependent variable and the independent variables are the offspring's environment and their parents' phenotypes. The regression coefficient on parents' phenotype equals heritability. With environmentally-altered meiotically-heritable epigenetic signals, we need to add a new independent variable to the regression: the environment of the parents [17]. Its regression coefficient equals the component of narrow-sense heritability due to environmentally-altered meiotically-heritable epigenetic signals.

If environment of the parents is omitted from the above parent—offspring regression, as is traditionally the case, then environmentally-altered meiotically-heritable epigenetic effects will be erroneously ascribed to other regression coefficients, especially genotype-by-environment interactions [17]. Sib analysis, which is the other standard quantitative genetic method, is not available for environmentally-alterable meiotically-heritable epigenetic signals because it does not provide a method for including environment of the parents.

Ramifications for neo-Lamarckian medicine

The above procedure for estimating meiotically-heritable epigenetic signals places neo-Lamarckian evolution on firm theoretical ground. Neo-Lamarckian evolution describes environmental variables altering heritable signals, an idea that even Darwin espoused [31,32]. Meiotically-heritable epigenetics corroborates this notion, albeit in a more random and less directed fashion than Lamarck had proposed [33,34]. Yet, despite neo-Lamarckian effects having first been identified long ago [26,35], neo-Lamarckian evolution is usually caricatured as a mode of evolution that was completely debunked in the mid 1800s (however, see [6]). Greater theoretical and empirical understanding of environmentally-altered meiotically-heritable epigenetic signals will help in correctly ascribing components of heritability. This is essential in so much of medicine, where there is great emphasis on discerning whether diseases are caused by genetics or the environment. Environmentally-altered meiotically-transmitted epigenetic effects show that this is a false dichotomy. In neo-Lamarckian medicine, the environment can heritably alter the epigenome.

Neo-Lamarckian medicine could have enormous public health and environmental policy implications. If industrial contaminants adversely affect organisms (including humans) by meiotically-heritably altering their epigenetic signals, then cleaning up these contaminants will not necessarily remove the problem. Once contaminants have adversely altered an organism's epigenetic signals, this harm will be transmitted to future generations even if subsequent generations are not exposed to the contaminant. Such effects have been documented for carcinogens acting by altering epigenetic signals [36,37]. This provides much greater impetus on preventing and correcting environmental contamination than currently exists. Neo-Lamarckian medicine changes the focus from cure to prevention [38].

The role of meiotically-heritable epigenetic signals in preventing and triggering cancers is intuitively plausible. Epigenetic signals, especially cytosine methylation, are known to regulate gene activity [14]. In particular, methylation of promoter genes almost always suppresses transcription of their structural gene. For many loci, downregulation of the gene is directly proportional to the amount of cytosine methylation on the promoter [18]. Environmental shocks are known to alter epigenetic signals, e.g. demethylate a locus (see [39]; in many of these instances, maintenance methylation following mitosis is precluded, rather than methylation being stripped away). Exposure to environmental shocks such as free radicals or other carcinogens generally alter cytosine methylation patterns on regulatory genes. If enough methylation is removed (including from mitotically produced daughter cells), gene activity can

become highly upregulated [40]. Because cell differentiation is largely the differential suppression of various gene functions, cells with substantial gene upregulation can become dedifferentiated and begin rapidly proliferating. This provides a potential route by which cancers could be triggered [41].

An alternative hypothesis is that tumour suppressor genes can be inactivated by too much methylation [41—43]. That is, methylation suppresses (downregulates) production of substances that preclude tumour growth, such as proteins that mop up free radicals. This alternative hypothesis is consistent with the observation that cancer cell telomeres are highly methylated [44]. Too much methylation is also associated with specific loci in certain tumours and these methylation levels can be environmentally-altered [45]. However, it is not known whether these specific epigenetic signals are meiotically-heritable.

Although I have focused on cancers, other diseases are believed to be of epigenetic origin, usually caused by aberrant levels of cytosine methylation. There is evidence that some forms of diabetes are caused by meiotically-heritable epigenetic effects [46]. Complex diseases are usually believed (and sometimes defined) to be caused by multiple genes and multiple environmental factors [47]. But complex diseases could just as readily be caused by a single gene on which is attached a variety of possible meiotically-heritable epigenetic signals at a single genetic locus [48], i.e. one locus with several epialleles. This is particularly plausible because there are usually multiple epigenetic layers on top of a single nucleotide locus (e.g. cytosine methylation, heterochromatin, histone acetylation). Although there does not yet appear to be any definitive evidence linking meiotically-heritable epigenetic signals with complex diseases, their strong genotype-by-environment interactions [49] makes them ripe for being studied in light of neo-Lamarckian medicine.

There have been a few attempts at applying Darwinian medicine to diseases caused by molecular epigenetic defects, such as cytosine methylation errors in Prader—Willi syndrome patients [50]. But this disease is probably caused by a meiotically-heritable epigenetic defect, rather than a genetic defect [51]. In fact, there does not appear to be any literature on evolutionary medicine being applied to diseases with meiotically-heritable epigenetic aetiology. My hope in introducing this conceptual foundation for neo-Lamarckian medicine is that those working in human medical genetics will begin considering meiotically-heritable epigenetic effects as possible causes for disease. If such diseases are found to be triggered by environmental perturbations, such as carcinogens, then there need to be concerted efforts made to prevent such diseases before subsequent generations are adversely affected.

Conclusion

Environment in one generation can affect subsequent generations by altering epigenetic signatures. If environmentally-altered meiotically-heritable epigenetic effects are widespread (which is an important open empirical question), they have the potential to alter paradigmatic views of evolutionary medicine, including abjuring the nature versus nurture dichotomy that pervades so much of modern medicine. With these effects, we may find the molecular basis for genotype-by-environment interactions. Their application in evolutionary medicine, called neo-Lamarckian medicine, could shift emphasis from cure to prevention of diseases.

References

[1] Williams GC, Nesse RM. The dawn of Darwinian medicine. Quart Rev Biol 1991;66:1—22.
[2] Stearns SC, Ebert D. Evolution in health and disease: work in progress. Quart Rev Biol 2001;76:417—32.
[3] Nesse RM. Culture and medicine: How is Darwinian medicine useful? West J Med 2001;174:358—60.
[4] Jablonka E, Lamb MJ. Epigenetic inheritance in evolution. J Evol Biol 1998;11:159—83.
[5] Jablonka E, Lamb MJ. The inheritance of acquired epigenetic variations. J Theor Biol 1989;139:69—83.
[6] Jablonka E, Lamb MJ. Epigenetic inheritance and evolution: the Lamarckian dimension. Oxford: Oxford University Press; 1995.
[7] Waddington CH. Organizers and genes. Cambridge: Cambridge University Press; 1940.
[8] Hall BK. Evolutiuonary developmental biology. London: Chapman & Hall; 1992.
[9] Caspersson T, Schultz J. Nucleic acid metabolism of the chromosomes in relation to gene reproduction. Proc Natl Acad Sci USA 1940;26:507—15.
[10] Brachet J. La détection histochimique des acides pentosenucléiques. Comptes Rend Soc Biol 1940;133:88—90.
[11] Fisher RA. The genetical theory of natural selection. Oxford: Clarendon Press; 1930.
[12] Crick F. Central dogma of molecular biology. Nature 1970;227:561—3.
[13] Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet 1975;14:9—11.
[14] Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science 1975;187:226—32.
[15] Holliday R. Epigenetics: an overview. Dev Genet 1994;15:453—7.

[16] Holliday R. The inheritance of epigenetic defects. Science 1987;238:163–70.
[17] Gorelick R. Conditional additive genetic variance, submitted for publication.
[18] Boyes J, Bird A. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. EMBO J 1992;11:327–33.
[19] Wolff GL, Kodell RL, Moore SR, Cooney CA. Maternal epigenetics and methyl supplements affect *agouti* gene expression in A^{vy}/a mice. FASEB J 1998;12:949–57.
[20] Rakyan VK, Chong S, Champ ME, et al. Transgenerational inheritance of epigenetic states at the murine $Axin^{Fu}$ allele occurs after maternal and paternal transmission. Proc Natl Acad Sci USA 2003;100:2538–43.
[21] Sano H, Kamada I, Youssefian S, Katsumi M, Wabiko H. A single treatment of rice seedlings with 5-azacytidine induces heritable dwarfism and undermethylation of genomic DNA. Mol Gen Genet 1990;220:441–7.
[22] Flavell RB, O'Dell M. Variation and inheritance of cytosine methylation patterns in wheat at the high molecular weight glutenin and ribosomal RNA gene loci. Development 1990;(1990 supplement):15–20.
[23] Ekwall K, Olsson T, Turner BM, Cranston G, Allshire RC. Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. Cell 1997;91:1021–32.
[24] Jollos V. Inherited changes produced by heat treatment in *Drosophila*. Genetica 1934;16:476–94.
[25] Wolffe AP, Matzke MA. Epigenetics: regulation through repression. Science 1999;286:481–6.
[26] Sturtevant AH. The effects of unequal cross-over at the Bar-locus in *Drosophila*. Genetics 1925;10:117–47.
[27] McClintock B. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA 1950;36:344–55.
[28] Crouse HV. The controlling element in sex chromosome behavior in *Sciara*. Genetics 1960;45:1429–43.
[29] Cowley DE, Atchley WR. Quantitative genetic models for development, epigenetic selection, and phenotypic evolution. Evolution 1992;46:495–518.
[30] Wolf JB, Frankino WA, Agrawal AF, Brodie ED, Moore AJ. Developmental interactions and the constituents of quantitative variation. Evolution 2001;55:232–45.
[31] Darwin CR. The origin of species by means of natural selection. 6th ed. London: John Murray; 1872.
[32] Darwin CR. The variation of animals and plants under domestication. 2nd ed. London: John Murray; 1875.
[33] Lamarck J-B. Systèm des animaux sans vertèbrates, ou tableau général...précéde du discours d'ouverture du cours de zoologie, donné dans le Muséum national d'Histoire Naturelle, Paris, 1801.
[34] Turner BM. Chromatin and gene regulation: mechanisms in epigenetics. Oxford: Blackwell Science; 2001.
[35] Baldwin JM. A new factor in evolution. Am Nat 1896;30: 441–51, 536–553.
[36] Li SF, Washburn KA, Moore R, et al. Developmental exposure to diethylstilbestrol elicits demethylation of estrogen-responsive lactoferrin gene in mouse uterus. Cancer Res 1997;57:4356–9.
[37] McLachlan JA, Burow M, Chiang TC, Li SF. Gene imprinting in developmental toxicology: a possible interface between physiology and pathology. Toxicol Lett 2001;120:161–4.
[38] Pollack R. Gene maps lead medicine down the wrong road. Perspect Biol Med 2002;45:43–5.
[39] Gorelick R, Osborne R. Inducing sex change and organogenesis from tissue culture in the endangered African cycad *Encephalartos woodii* (Cycadales, Zamiaceae). S Afr J Sci 2002;98:114–7.
[40] Jackson-Grusby L, Beard C, Possemato R, et al. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nature Genet 2001;27:31–9.
[41] Paulsen M, Ferguson-Smith AC. DNA methylation in genomic imprinting, development, and disease. J Pathol 2001;195:97–110.
[42] Herman JG, Latif F, Weng YK, et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc Natl Acad Sci USA 1994;91:9700–4.
[43] Agathanggelou A, Honorio S, Macartney DP, et al. Methylation associated inactivation of RASSF1A from region 3p21.3 in lung, breast and ovarian tumours. Oncogene 2001;20:1509–18.
[44] Tollefsbol TO, Andrews LG. Mechanisms for telomerase gene control in aging cells and tumorigenesis. Med Hypotheses 2001;56:630–7.
[45] Esteller M, Guo MZ, Moreno V, et al. Hypermethylation-associated inactivation of the cellular retinol-binding-protein 1 gene in human cancer. Cancer Res 2002;62: 5902–5.
[46] Bennett ST, Wilson AJ, Esposito L, et al. Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele. Nature Genet 1997;17:350–2.
[47] Cardon LR, Abecasis GR. Using haplotype blocks to map human complex trait loci. Trends Genet 2003;19:135–40.
[48] Metheny LJ, Cappione AJ, Skuse GR. Genetic and epigenetic mechanisms in the pathogenesis of neurofibromatosis type I. J Neuropathol Exp Neurol 1995;54:753–60.
[49] Strohman RC. Ancient genomes, wise bodies, unhealthy people: limits of a genetic paradigm in biology and medicine. Perspectives in Biology and Medicine 1993;37: 112–45.
[50] Haig D, Wharton R. Prader–Willi syndrome and the evolution of human childhood. Am J Hum Biol 2003;15: 320–9.
[51] Buiting K, Gross S, Lich C, Gillessen-Kaesbach G, El-Maarri O, Horsthemke B. Epimutations in Prader–Willi and Angelman syndromes: a molecular study of 136 patients with an imprinting defect. Am J Hum Genet 2003;72:571–7.

APPENDIX B

GENE DUPLICATIONS AND EVOLUTION OF SEX, PART 1,

EVOLUTION OF DIOECY AND SEX CHROMOSOMES

Biological Journal of the Linnean Society, 2003, **80**, 353–368. With 2 figures

Evolution of dioecy and sex chromosomes via methylation driving Muller's ratchet

ROOT GORELICK*

Department of Biology, Arizona State University, Tempe, AZ 85287–1501, USA

Received 13 October 2002; accepted for publication 7 April 2003

Why are there two sexes in certain species, instead of one hermaphroditic sex? Why are Y chromosomes shorter than X chromosomes, but only in certain lineages? I propose that differences between sexes are initially determined by differential methylation in nuclear DNA between females and males, driving Muller's ratchet. Methylation of promoters suppresses transcription, including loci coding for gamete production, thereby converting hermaphroditic individuals into females or males. Differential methylation of sex chromosomes suppresses recombination and increases mutation rate, thereby geometrically increasing the speed of Muller's ratchet. Higher mutability of methylated nucleotides plus loss of sex-determining function of previously methylated nucleotides provides selective pressure to excise these loci, resulting in shorter Y or W chromosomes. Derived lineages usually have more methylation than do ancestral ones, and hence have relatively shorter sex chromosomes. Methylation canalizes dioecy and degeneration of sex chromosomes. Latter stages of sex chromosome evolution may have occurred via other mechanisms, for example sexually antagonistic genes or chromosomal rearrangements. A few aberrant derived lineages lost most methylation, and their sex determination and sex chromosomes may have evolved via other means. Differential methylation provides a mechanism for early evolution of dioecy in anisogamous sexual diploid eukaryotes and of sex chromosomes in metazoans.

ADDITIONAL KEYWORDS: canalization – demethylation – differentially methylated region (DMR) – epigenetic – heterochromatin – temperature-dependent sex determination (TSD).

INTRODUCTION

For at least the past 3000 years, people have been speculating on the causes of sex and dioecy (Mittwoch, 2000). However, it took until the advent of modern genetics at the dawn of the 20th century for scientists to connect two distinct sexes with sex chromosomes (McClung, 1902; Stevens, 1905). For the next seven decades, insufficient molecular evidence existed for a more refined or reductionist model of evolution of dioecy. However, in the past quarter century, there has been an almost evangelical search for a single sex-determining locus on sex chromosomes (Wachtel & Tiersch, 1994; Mittwoch, 2000; but see Graves, 2002), ignoring substantial evidence that earlier researchers may have been correct that sex is determined by differential properties of whole sex chromosomes. My approach falls in between these two extremes; it can account for the evolution of dioecy via a single gene and via effects spread out more diffusely over multiple genes.

There is no single unified hypothesis explaining sex determination, not even when confining attention to metazoans or even to vertebrates. The problem is that environmental and chromosomal sex determination are usually considered to be different phenomena. I propose a hypothesis that explains both types of sex determination by focusing on epigenetic patterns across wide swaths of incipient sex chromosomes, rather than genetic patterns of nucleotides on one or a few loci. These epigenetic patterns consist of a binary code of methylation: is each nucleotide locus methylated or not? Methylation is the addition of a methyl group (CH_3) to a DNA nucleotide, usually to the 5′-carbon of cytosine (5-methylcytosine) in cytosine-guanine dinucleotides (CpG). Heterochromatic proteins bind to methylated nucleotides, causing the G-chromosomal banding patterns that are observable

*E-mail: cycad@asu.edu

under a light microscope (Miniou *et al*., 1997; Singer, Yordan & Martienssen, 2001; Richards & Elgin, 2002). Methylation patterns are highly heritable (Holliday, 1988; Vyskot *et al*., 1995). I propose that differential methylation of regulatory genes controlling production of primary sex characteristics, gonads, and/or gametes on a pair of homologous chromosomes provided the origin of two distinct sexes and proto-sex chromosomes.

When biologists think of sex chromosomes, they typically do not think of differential methylation of sex-determining loci. Instead, they think of a pair of chromosomes of unequal length in one of the two sexes. This distinction between dioecy and sex chromosomes is a result of chromosomes of unequal length being readily observable under a light microscope. Differential methylation levels are much less visible, although they are potentially visible as subtle differences in chromosomal banding patterns under a light microscope. Differential methylation of sex-determining loci is merely the first step in shortening of one of the two sex chromosomes (Solari, 1994 and Steinemann & Steinemann, 1998 point out that differential heterochromatin is the first step in forming heteromorphic sex chromosomes in some species, but fail to mention that heterochromatin often contains methylated DNA sequences).

The most widely accepted explanation for the evolution of sex chromosomes of unequal length is Muller's ratchet for the accumulation of deleterious mutations on one of the two sex chromosomes (Muller, 1914; Muller, 1964; Frota-Pessoa & Aratangy, 1968; Nei, 1970; Felsenstein, 1974). The problem with Muller's ratchet and other proposed models is in starting the process of sex chromosome evolution without invoking seemingly ad hoc assumptions, such as a mechanism by which recombination is initially suppressed. I modify this traditional model by using the effects of methylation to adjust two key parameters in Muller's ratchet: recombination and mutation rates. This accelerates the rate of Muller's ratchet, providing a rationale for why there exists only one pair of sex chromosomes, and further explaining the close relationship between evolution of dioecy and sex chromosomes.

There probably exists a plurality of mechanisms for evolution of dioecy and sex determination (West, Lively & Read, 1999). My proposal tries to elucidate the most widespread and likely origin. In plants, where there is much more movement of genes between nuclei and organelles (Palmer *et al*., 2000), there is a much higher likelihood of cytoplasmic sex determination. In those few aberrant and highly derived animal and fungal lineages in which cytosine methylation was lost, dioecy and sex chromosomes may have evolved via mechanisms not associated with methylation. Unfortunately, these aberrant derived lineages contain those taxa that are most often studied to determine the origins of dioecy and sex chromosomes: the yeasts *Saccharomyces cervisiae* and *Schizosaccharomyces pombe*, the fruit fly family Drosophilidae, and the nematode *Caenorhabditis elegans* (Riggs & Pfeifer, 1992; Wolffe & Matzke, 1999; Gowher, Leismann & Jeltsch, 2000; Lyko, Ramsahoye & Jaenisch, 2000). *Drosophila* have highly derived and atypically complicated sex-determining systems (Dübendorfer *et al*., 2002), hence extensive earlier work on the evolution of sex determination and sex chromosomes derived from studies of *Drosophila* are probably not applicable to most other taxa. The evolution of sex determination is also well studied in mammals, but here the sex-determining mechanisms are probably so highly canalized that it is impossible to gain any glimpses of the early evolution of sex determination. Therefore, our best windows into the evolution of dioecy and sex chromosomes come from relatively ancestral animal taxa.

The first half of this paper provides this theory of methylation driving the evolution of dioecy and sex chromosomes, while the second half provides several testable predictions and details regarding testing.

THEORY

Evolution of dioecy

It is generally believed that two distinct sexes evolved via a single sex-determining locus. However, this fails to explain how such a locus could have independently evolved many times, why chromosomal banding patterns differ over large portions of sex chromosomes and not just near the putative sex-determining locus, how recombination between incipient sex chromosomes is initially suppressed, and the mechanisms by which environmental and hormonal sex determination operate. To rectify this, I propose an alternative model of evolution of dioecy via differential methylation of sex-controlling portions of the genome. Differential methylation could be on a single gene, which may then be interpreted as a sex-determining locus, but could just as well occur on multiple unlinked genes.

Methylation (CH_3) is invariably attached symmetrically on double-stranded DNA at specific palindromic nucleotide sequences. This symmetry allows for faithful copying of methylation patterns during chromosomal replication, in which maintenance methylation corrects asymmetries by adding methylation to the unmethylated strand (Woodcock *et al*., 1997). Methyl groups are usually attached to the 5′-carbon of cytosine (5-methylcytosine) in CpG dinucleotides, or less frequently to CpNpG trinucleotides, where N can be any nucleotide (Ramsahoye *et al*., 2000). I disregard the other forms of methylation, N^6-methylade-

nine and N^4-methylcytosine, because they are absent from all meiotic genomes. Methylation patterns are highly heritable and form a code akin to and on top of nucleotide sequences, but with a different alphabet. Methyl groups attach directly to DNA nucleotides, hence they are referred to as epigenetic phenomena ('epi-' means 'on top of').

I hypothesize that dioecy and sex chromosomes originated in ancestral diploid hermaphrodites as a pair of ordinary homologous chromosomes (autosomes) in which one chromosome had more methyl groups attached to a sex-controlling region than did its homologue (see de Almeida-Toledo *et al*., 2001 for a similar notion). The chromosome with the more highly methylated sex-controlling region was the incipient Y or W chromosome. Thus, a second distinct sex was initially determined solely by methylation patterns. Moreover, genomic imprints (i.e. parent-specific 'differentially methylated regions') are the cause rather than effect of having two sexes. Below I show why sex-specific methylation and proto-sex chromosomes were initially confined to one pair of homologous autosomes, consistent with theoretical models and empirical data from all eukaryotes that have multicellular diploid stages (Hurst & Hamilton, 1992).

Methylation blocks binding sites for enzymes that mediate transcription (Iguchi-Ariga & Schaffner, 1989; Tate & Bird, 1993; Henry *et al*., 1999; Yung *et al*., 2001) (Fig. 1) in several interrelated ways (see Gorelick, 2003 and references therein). Transcription is invariably suppressed if promoters are methylated, especially in dense regions of cytosine-guanine dinucleotides (Futscher *et al*., 2002). In some but not all promoters, the degree of suppression of transcription is proportional to the density of cytosines that are methylated (Boyes & Bird, 1992; Tate & Bird, 1993).

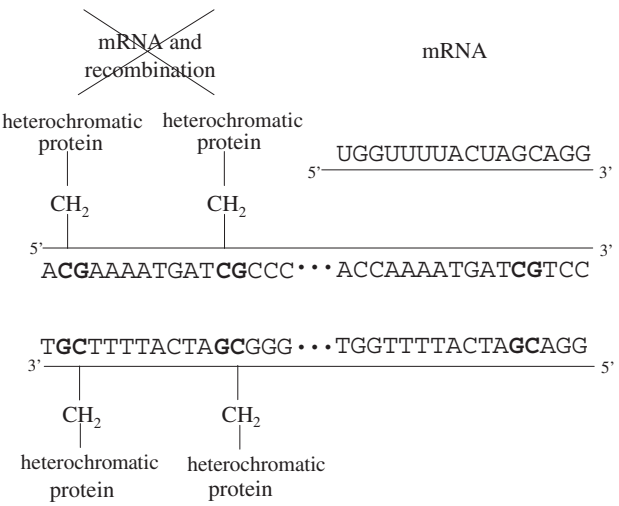

Figure 1. Effects of methylation on recombination and transcription.

In hermaphroditic anisogamous diploid eukaryotes, I hypothesize that differential methylation of a locus whose gene products are responsible for primary sex characteristics or gamete formation caused the origin of a second sex. Methylation regulates production of gene products in a regulatory cascade that produces female or male sexual characteristics (Iannello *et al*., 1997; Grant, 1999; Iannello *et al*., 2000; Griswold & Kim, 2001). Suppression of such gene products could suppress the production of either eggs or sperm (Charlesworth & Charlesworth, 1978). Whichever pair of autosomes this differential methylation first occurred on would become the proto-sex chromosomes. If the differential methylation suppressed production of male (female) sex hormones or some other gene product that contributes to production of sperm (egg) cells, then this would become the proto-Y (W) chromosome. A second pair of autosomes could not subsequently form another set of proto-sex chromosomes because they could only do so by indirectly regulating the production of eggs or sperm, which would either result in (1) no effect if both methylated sites affected production of the same form of gamete or (2) zero fitness if both methylated sites inhibited production of different forms of gametes. Differential methylation could affect regulation of genes on other chromosomes, however, so long as these genes do not affect sex determination and production of gametes. Furthermore, the chromosomes on which these other genes reside will not have loci excised. Below, I further discuss why there can only be one pair of homologous chromosomes of unequal length.

'In principle, one genetic or epigenetic signal . . . would suffice for sexual differentiation' (Dübendorfer *et al*., 2002: 75). Only a small regulatory change, such as differential methylation, would be required to create the first female or male individual in a hermaphroditic species. In hermaphroditic individuals, female (male) functions are suppressed in male (female) sex organs. Therefore, a strictly female individual could be created via co-option (by the male sex organs) of the regulatory machinery that originally suppressed male function in female sex organs of hermaphrodites. For ancestral lineages, sex determination was probably controlled by one or a few loci. However, with repeated duplications of these genes over evolutionary time, derived lineages may have multiple loci involved in sex determination.

If sex is determined by many independent (but possibly paralogous) genes and each locus requires relatively dense promoter methylation to be suppressed, then the old notion of global determination of sex is correct. If sex is determined by one or a few genes whose resulting biochemical pathways are tightly

interlinked, then the modern notion of a single sex-determining locus is correct (e.g. Canning & Lovell-Badge, 2002), especially if these gene products have promoters that are down-regulated by the addition of only one or a few methyl groups.

Evolution of sex chromosomes

Evolution of sex chromosomes is usually explained by a traditional application of Muller's ratchet. Muller's ratchet is a population genetic model of how the minimum number of point mutations (hereinafter mutations) per individual accumulates over evolutionary time in a finite population with reduced recombination, i.e. with reduced crossing over between distinct homologues. The class of individuals with the minimum number of mutations is lost via genetic drift and mutation. The speed of Muller's ratchet is proportional to the per-genome mutation rate (i.e. per-locus mutation rate times genome length) and inversely proportional to the effective population size and recombination rate (Nei, 1970; Haigh, 1978). Muller's ratchet only applies when the haploid stage of the life cycle is immune from selection. Technically, Nei's (1970) model only accounts for a single iteration of the ratchet. An extension of Nei's model to multiple steps of Muller's ratchet does not exist for sexual lineages. I therefore rely on his single step approach, but refer to the output of his model as speed of Muller's ratchet.

Evolution of sex chromosomes has also been explained using other forms of the Fisher–Muller model (Felsenstein, 1988), such as genetic hitchhiking or Hill–Robertson effect, and by retrotransposon traps. None of these models explain how recombination is initially suppressed in the vicinity of putative sex-determining loci, especially in genetically identical sex chromosomes, although this is a prerequisite for each model. Nor do these models explain why derived lineages have more highly degenerated sex chromosomes compared with ancestral lineages nor why derived lineages usually have two genetically distinct sexes.

I propose that methylation suppresses recombination, thereby starting Muller's ratchet, that methylation increases mutation rate and decreases recombination rate, thereby speeding up Muller's ratchet, and that mutation and epimutations of methylated nucleotides provide the selective force for deletion of loci from Y and W chromosomes.

The traditional Muller's ratchet model has recombination rate and per genome point mutation rate given exogenously. Nei (1970) appears to have provided the only mathematical model of Muller's ratchet that explicitly includes recombination rate as a parameter. The speed of Muller's ratchet is an increasing function of per-genome mutation rate and a decreasing function of recombination rate (Nei, 1970). Recombination and mutation rates only appear as products of each other in Nei's formulation of Muller's ratchet. In the remainder of this section, I show that recombination rates are inversely proportional to methylation level and that mutation rates are directly proportional to methylation level. Therefore, an increase in methylation level increases the speed of Muller's ratchet in two different ways: by suppressing recombination and by increasing mutation. Since these two factors appear as products of each other in Nei's formulation, this results in a geometric increase in the speed of Muller's ratchet. Once recombination and mutation rates are considered functions of methylation level, methylation both starts and accelerates Muller's ratchet.

Methylation suppresses recombination (Holliday, 1984; Holliday, 1988; Colot & Rossignol, 1999) in the same way it suppresses transcription. Heterochromatic proteins bind to methylated loci (Miniou *et al*., 1997; Singer *et al*., 2001; Richards & Elgin, 2002), thereby blocking sites for enzymes that mediate recombination (Catcheside, 1986; Hsieh, Meyn & Camerini-Otero, 1986; Rauth *et al*., 1986). It is currently believed that methylation suppresses recombination of any loci that are known recombination sites. What is important here is starting Muller's ratchet, so there simply need to be a few sites that have suppressed recombination, such as promoters that control sex determination.

Contrary to many claims, sexually antagonistic genes (Fisher, 1931; Bull, 1983; Rice, 1987a, 1996) do not in fact provide a mechanism for initially suppressing recombination. They only accelerate already existing local suppression of recombination. The sexually antagonistic genes model contains two assumptions: (1) there is a single or there are a few sex-determining loci and (2) genes that locally suppress recombination are common near those sex-determining loci (Bull, 1983; Rice, 1996). With these two strong assumptions, sexually antagonistic genes can cause the degeneration of Y or W chromosomes and may play a role in the latter stages of sex chromosome evolution, but probably played no role in incipient sex chromosome formation. Differential methylation obviates these two assumptions and does so by providing the molecular mechanisms for initially suppressing recombination, mechanism that the leading supporters of the sexually antagonistic genes hypothesis claim are unknown (e.g. Rice, 1996: 334).

Methylated cytosine mutates to thymine at a much higher rate than does unmethylated cytosine. I review the deamination pathways by which methylated and unmethylated cytosine incur point mutations and then discuss correction mechanisms by which these mutations are often repaired (Fig. 2).

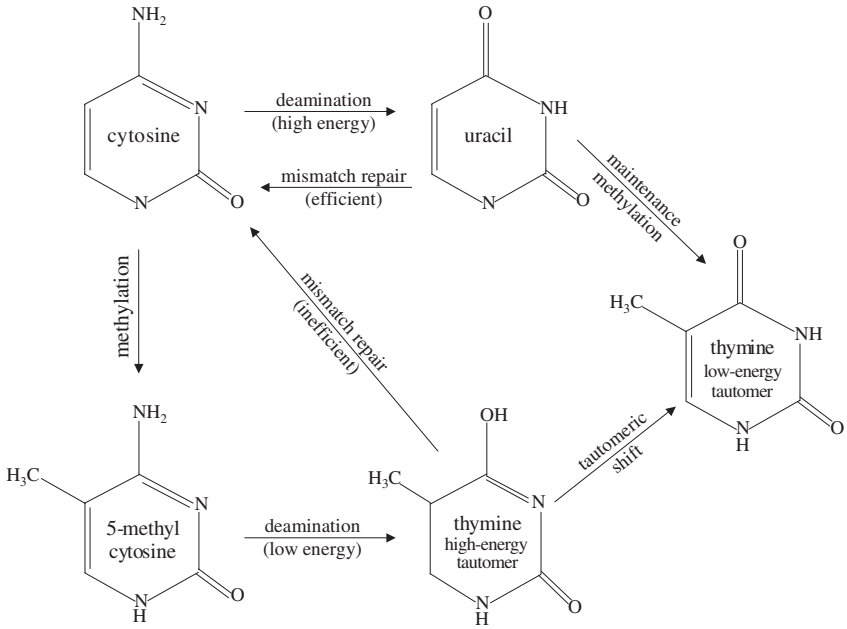

Figure 2. Pathways for methylation, point mutation, and mismatch repair. High-energy tautomer of thymine is 4-hydroxy-5-methyl-2-pyrimidinone. Low-energy tautomer of thymine is 5-methyl-2,4-pyrimidinedione. Methylation of cytosine can be via de novo methylation.

Both methylated and unmethylated cytosine nucleotides incur point mutations (transitions) via deamination. Point mutations of unmethylated cytosine to thymine are via hydrolytic or enzymatic deamination to uracil, which – if not corrected by subsequent mismatch repair – are converted by maintenance methylation to the commonest and lowest energy form of thymine, the dioxo-tautomer, 5-methyl-2,4-pyrimidinedione (Yang, Jones & Shibata, 1996). Point mutations of 5-methylcytosine to thymine are via hydrolytic deamination to a monoxo-tautomeric form of thymine, 4-hydroxy-5-methyl-2-pyrimidinone, which – if not corrected by subsequent mismatch repair – are almost immediately converted to the commonest and lowest energy form of thymine, the dioxo-tautomer, 5-methyl-2,4-pyrimidinedione (Norberg & Vihinen, 2001). Most likely, point mutations of methylated or unmethylated cytosine are spontaneous and methylation of cytosine increases the mutation rate because the attached methyl group lowers the energy required for deamination (Shen, Rideout & Jones, 1994; Yang *et al*., 1996). The energy required for deamination differs if one looks at single oligonucleotides, single-stranded DNA, or double-stranded DNA (Yang *et al*., 1996), indicating that the conformational changes resulting from DNA methylation probably also affect deamination and mutation rates of methylated cytosine. Alternatively, housekeeping enzymes that bind to methylated cytosine, but depend on the structure of DNA molecules, may trigger this deamination (Norberg & Vihinen, 2001). Regardless of the details, methylated cytosine has approximately three times the deamination rate of unmethylated cytosine, and hence a higher mutation rate.

Relative rates of point mutation of methylated and unmethylated cytosine to thymine are also affected by mismatch repair mechanisms that bind to the intermediate products of deamination (Fig. 2). Mismatch repair enzymes appear to detect hydrogen bonding between the incumbent uracil or 4-hydroxy-5-methyl-2-pyrimidinone and the unmethylated or methylated cytosine's original complement of guanine (Barrett *et al*., 1998). Mismatch repair enzymes then convert uracil or 4-hydroxy-5-methyl-2-pyrimidinone to unmethylated cytosine. Mismatch repair does not correct all such point mutations because the mismatch repair enzymes sometimes operate after DNA replication, hence there is no complementary template to use in correcting errors (Brown & Jiricny, 1987; Jones *et al*., 1992). Mismatch repair involves a different, but probably homologous, set of enzymes for point mutations of methylated vs. unmethylated cytosine (Barrett *et al*., 1998). Mismatch repair of hydrogen bonded uracil : guanine pairs is six thousand times more efficient than that of 4-hydroxy-5-methyl-2-pyrimidinone : guanine pairs (Schmutte *et al*., 1995). Heuristically, this is entirely plausible because uracil is recognized as a foreign nucleotide in double-stranded nuclear or mitochondrial DNA (uracil is never used for coding in double-stranded DNA in any

organism), whereas 4-hydroxy-5-methyl-2-pyrimidinone can easily go undetected by mismatch repair enzymes because it is merely a tautomer of thymine. Because it is a low-energy reaction, tautomeric conversion of 4-hydroxy-5-methyl-2-pyrimidinone to 5-methyl-2,4-pyrimidinedione occurs much more rapidly than does maintenance methylation of uracil to thymine, thereby precluding the opportunity for mismatch repair of many deaminated molecules of 5-methylcytosine. Therefore, the ultimate genetic point mutation rate – following deamination, maintenance methylation, and mismatch repair – is almost 20 000 times greater for methylated cytosine . . . so much so that methylated cytosines cause 30–40% of all germ-line point mutations in humans (Jones *et al*., 1992).

If a methylated cytosine is converted to unmethylated cytosine via deamination and mismatch repair, that individual retains an altered methylation signature. From a genetic perspective, this is not a problem because both methylated and unmethylated cytosine are transcribed into identical messenger RNA, if they are transcribed at all (Jones *et al*., 1992). However, from an epigenetic perspective, this can have dire consequences, especially if it occurs at a sex-determining locus. The only opportunity for correcting this loss of methylation is during global demethylation and de novo methylation of the meiotic (nuclear) genome at or immediately following gamete formation and syngamy. However, this is too late for an individual that depends on its methylation to regulate production of primary sexual characters and gametes.

Increased mutation rate of methylated (vs. unmethylated) cytosine into thymine increases the speed of Muller's ratchet. Once genes containing sex-determining methylation patterns contain nucleotides that mutate from methylated cytosine into thymine, there is strong selective pressure to excise these loci, resulting in immediate degeneration of the proto-Y chromosome into a true (i.e. shorter) Y chromosome. Note that the only way in which one of two differentially methylated homologous chromosomes will initially have loci excised is if the differential methylation suppressed production of eggs or sperm. Because it is dire to completely suppress production of both eggs and sperm and of no consequence to suppress two separate parts of the same regulatory pathway that produces either eggs or sperm, only one set of autosomes can have loci initially excised and thereby become sex chromosomes in any given lineage. Existence of only a single pair of sex chromosomes is not predicted by any other theory, including those hypothesizing a single putative sex-determining locus, because such putative loci can be duplicated.

As with all previous theories regarding degeneration of sex chromosomes (e.g. Charlesworth, 1991; Rice, 1994, 1996; Charlesworth & Charlesworth, 2000), I remain mute about the precise mechanism by which mutated and epimutated loci are excised from non-recombining portions of sex chromosomes. Although such loci may be responsible for severe fitness disadvantages, this does not explain how they are excised. Because recombination is locally suppressed in all of these models, excision cannot be due to increased crossing over. However, with methylation driving evolution of sex chromosomes, a lack of crossing over provides an absolute fitness disadvantage insofar as X or Z chromosomes do not accumulate beneficial mutations more rapidly than do Y or W chromosomes, as would be case with background trapping (Rice, 1996).

Once recombination is initially suppressed and Muller's ratchet has started on a pair of proto-sex chromosomes, there is no longer a requirement that the mutations driving the ratchet be associated with sex determination. Mutated loci merely need to be tightly linked to an ancestral sex-determining locus, i.e. be on the same chromosome with suppressed recombination. Thus, ceteris paribus, when the heterogametic proto-sex chromosome has more methylation of any loci, it will have a faster Muller's ratchet.

Although I propose that methylation is largely responsible for the origin of two distinct sexes and sex chromosomes, other mechanisms could be responsible for the latter stages in this evolution. Inversions and translocations are responsible for many recently derived structural differences in sex chromosomes. For example, the relative sizes of sex chromosomes across all mammals and across all species in the lizard genus *Sceloporus* are probably due to chromosomal rearrangements, and not to gradual deterioration of a Y or W chromosome (Sites *et al*., 1992; Graves, 1995b).

Because sex is initially determined by differential methylation and (evolutionarily) later by gene deletions from Y or W chromosomes, differences between proto-sex chromosomes canalize dioecy. Originally, each pair of homologous chromosomes had the same lengths and the same methylation patterns, and these lineages were cosexual. Differential methylation of one pair of homologous chromosomes canalized dioecy (i.e. loss of hermaphrodites) in diploid eukaryote lineages, although these lineages may still have been cosexual as with temperature-dependent sex determination. With sufficient differences in methylation between the two sexes, sex change became unidirectional or nonexistent and heterogameity became canalized. Evolution from X/Y to Z/W, or vice versa, became impossible. Furthermore, dioecy itself became canalized. Charles Darwin was the first to suggest that the sex of animals and plants is canalized from an initial hermaphroditic condition (Darwin, 1873; Stauffer, 1975). However, lack of infor-

mation regarding heredity precluded him from hypothesizing about molecular or physiological causes for this canalization.

The most likely reason why nobody as yet has independently applied methylation to Muller's ratchet for the evolution of sex chromosomes is that virtually all modern biologists working on this problem have used *Drosophila* as their study organism (Charlesworth, 1978; Rice, 1987b; Charlesworth, 1991; Steinemann & Steinemann, 1992; Rice, 1996; Charlesworth & Charlesworth, 2000). The family Drosophilidae and other isolated lineages of classical model organisms, such as the nematode *Caenorhabditis elegans* and the yeast *Saccharomyces cervisiae*, independently lost all or most DNA methylation in the recent evolutionary past, while 'normal' higher levels of methylation were retained in all related lineages (Urieli-Shoval *et al*., 1982; Proffitt *et al*., 1984; Riggs & Pfeifer, 1992; Wolffe & Matzke, 1999; Gowher *et al*., 2000; Lyko *et al*., 2000). This highly derived lack of (most) methylation deprived researchers working with *Drosophila* of the key mechanisms by which the sex of an individual is determined, recombination is suppressed, and mutation rate is increased.

Although the arguments in this section on the evolution of sex chromosomes appear to apply to all eukaryotes, they in fact only apply to metazoans because of an underlying assumption behind Muller's ratchet. One of the predictions below posits that other eukaryotes, such as fungi and plants, are immune from the effects of Muller's ratchet due to their extensive gene expression during haploid stages. In contrast, my arguments regarding evolution of dioecy apply to all anisogamous sexual diploid eukaryotes. Their nuclear genomes can become differentially methylated, even though Muller's ratchet cannot subsequently operate on them.

PREDICTIONS

I provide testable predictions arising from the above theory, including several corollaries (Table 1). I also give details regarding the testing of these predictions.

Dioecious lineages have heteromorphic sex chromosomes

I predict that all dioecious lineages will have at least slightly different sex chromosomes. If the homologous chromosomes have identical nucleotide sequences, then they will have different methylation patterns. These chromosomes will have differences in G-banding patterns because heterochromatic proteins bind to methylated DNA (Schmid & Haaf, 1989).

Only after comprehensive data on methylation levels becomes available will it be possible to test whether dioecious lineages with putative sex chromosomes of equal length have differential methylation levels between the two homologues in the heterogametic sex. The output of this analysis would be a phylogenetic comparison of dioecy/monoecy (or dioecy/cosexual) vs. extent of methylation difference between the sexes or between the two proto-sex chromosomes. Cosexual lineages are defined as those in which there exist some individuals that are capable of producing both eggs and sperm or have environmentally determined sex, i.e. lack of strict dioecy (Lloyd, 1980). Comparisons of methylation differences will be especially important if they are of promoters of genes known to code for production of gametes or primary sexual characters. This analysis may require a separate comparison for each large taxonomic group, such as vertebrates and flowering plants. The most appropriate tool would be bisulphite sequencing of promoters affecting sex determination, especially using tools that assess density of methylation (Galm *et al*., 2002; Gitan *et al*., 2002). A simpler approach would be to use reverse-phase high-performance liquid chromatography (HPLC). However, the results would be equivocal because the relevant promoters undoubtedly occupy a small fraction of the genome, which may not have methylation patterns similar to the rest of the genome.

A corollary is that cosexual lineages will contain a pair of sex chromosomes with virtually the same methylation patterns and the same lengths. A perusal of the literature corroborates this corollary, especially for vertebrates, for which the nature of sex chromosomes has been encyclopedically documented (Solari, 1994), assuming that G-banding reflects relative methylation levels.

Sex change can be induced in lineages with equal length sex chromosomes

I predict that most environmentally and hormonally induced sex changes in flowering plants and vertebrates occur via demethylation and these methylation changes are heritable (LoSchiavo *et al*., 1989; Arnholdt-Schmitt *et al*., 1991; Dorazi, Chesnel & Dournan, 1995; Demeulemeester, Van Stallen & De Proft, 1999; Murphy & Jirtle, 2000; Tatra *et al*., 2000). Recent work indicates that sex hormones determine sex via heritable changes in methylation patterns (McLachlan, 2001). Therefore, in lineages in which sex is solely determined by small differences in methylation between proto-X and Y (or Z and W) chromosomes, I predict that it should be possible to change the sex of an individual by subjecting it to demethylating compounds (Vyskot *et al*., 1995). If this theory is correct, these sex changes should primarily be back to the ancestral hermaphroditic condition.

Testing can be done with any dioecious taxon in

Table 1. Predictions arising from methylation driving evolution of dioecy and sex chromosomes

Prediction	Data required to test	Tentative results
Dioecy yields heteromorphic sex chromosomes	Bisulphite sequences; phylogeny	None
Cosexual lineages have both sex chromosomes with virtually the same methylation patterns and lengths	Bisulphite sequences, HPLC, or G-chromosomal banding	Confirmed, based on G-banding (Solari, 1994)
In lineages with equal length sex chromosomes, sex change can be induced	Results from application of 5-azacytidine, ethionine or 5-methyl-deoxcytidine	None
If sex chromosomes are of different lengths, then sex change cannot be induced by altering methylation levels	Results from application of 5-azacytidine, ethionine or 5-methyl-deoxcytidine	None
Sex change is unidirectional, to sex with less methylation of sex-determining loci	Bisulphite sequences, HPLC, or G-chromosomal banding	None
If both sexes have equal amounts but different distributions of methylation of sex-determining loci, then changes in sex can be in either direction, depending on which methyl groups are removed	Bisulphite sequencing	None
Derived lineages have shorter Y (or W) chromosomes	Genome length, phylogeny	Tentative corroboration in vertebrates (Solari, 1994)
Different length sex chromosomes yields dioecy	Chromosome length	Tentative corroboration (Solari, 1994)
Diploid cosexual lineages cannot be descended from dioecious lineages	Phylogenies accurately showing transitions between cosexuality and dioecy	No known counterexamples
Allopolyploidy provides an escape from canalization of dioecy	Whether polyploidy was a result of alloploidy or autoploidy; gene phylogenies with transitions between cosexuality and dioecy	No known counterexamples, but most data necessary for testing is nonexistent
TSD requires homomorphic sex chromosomes	Chromosome lengths	Tentatively corroborated in vertebrates (Olmo, 1986; Spotila *et al*., 1994)
Highly heteromorphic sex chromosomes imply dioecy, hence GSD	Chromosome lengths	Tentatively corroborated in vertebrates (Solari, 1994)
Degeneration of sex chromosomes will be rapid	Estimates of evolutionary rates of sex chromosome degeneration	None
Plant and fungal gametophytes express most genes expressed by their diploid stages	Microarray comparisons of haploid and diploid stages	None
Substantial genetic drift implies higher likelihood of cosexuality	Sister taxa having planktonic and non-planktonic larvae; sex chromosome lengths	None
Most marine invertebrates have identical sex chromosomes and are thus hermaphroditic	Knowledge of cytology and breeding systems	Seems to be true
Terrestrial and marine invertebrates without ancestral planktonic larvae have distinctive sex chromosomes and are strictly dioecious	Knowledge of cytology and breeding systems	Seems to be true, but this may be biased by insects and benthic nematodes

GSD, genetic sex determination; HPLC, high-performance liquid chromatography; TSD, temperature-dependent sex determination.

which there is no difference in length between sex chromosomes and sex is determined solely by small differences in methylation levels. Dosages of 5-azacytidine or ethionine could be applied so that the resulting methylation level corresponds with that of the less methylated sex (Gorelick & Osborne, 2002), realizing that the 'wrong' methyl groups might get stripped away (Giancotti *et al*., 1995). Application of 5-azacytidine to males of the dioecious flowering plant *Melandrium album* changed some of them into androhermaphroditic indi-

viduals, but it had no effect when applied to female plants (Janousek, Siroky & Vyskot, 1996). This indicates that methylation had a female-suppressing function in male individuals. It may also be possible to add methylation to the less methylated sex by applying 5-methyl-deoxcytidine (Holliday & Ho, 1991; Nyce, 1991; Holliday & Ho, 1995).

A corollary is that if sex chromosomes are of different lengths, then sex change producing sexually fertile individuals cannot be induced by altering methylation levels or by any other method aimed at single genes (Solari, 1994).

Sex change is unidirectional, to the sex with less methylation of sex-determining loci

I predict that sex changes will almost always be in one direction – from the more to the less methylated sex or to a relatively unmethylated hermaphrodite – when methylation changes are only on sex-determining promoters. The motivation behind this prediction is that it is much simpler (i.e. requires much less complicated biochemical machinery) to remove methyl groups than it is to add them (in many of these instances, maintenance methylation following mitosis is precluded, rather than methylation being stripped away). Analogously, it is much simpler to teach someone to erase a chalkboard than to teach them to write something intelligent on one (an entropy argument). It is well known that sex change is usually only in one direction (Nakashima, Kuwamura & Yogo, 1995; Pannell, 1997; Kuwamura & Nakashima, 1998). What is new here is that I propose that the change is almost always to the less methylated sex-determining loci.

It is possible to have differential methylation of females and males with both sexes having equal amounts of methylation of promoters or of any and all portions of their genomes (Pardo-Manuel de Villena, de la Casa-Esperón & Sapienza, 2000). In this case, I predict that changes in sex can be in either direction, simply depending on which methyl groups are removed.

This prediction appears to be corroborated. Sex change is usually only in one direction (Nakashima *et al*., 1995; Pannell, 1997; Kuwamura & Nakashima, 1998). In the rare instances where sex change can occur in both directions in a lineage, it would be helpful to determine whether the methylation levels between the two sexes are virtually identical, at least on those promoters that control for gamete production and primary sexual characters.

Derived lineages have shorter Y (or W) chromosomes

Derived lineages often evolved via duplication of regulatory genes, accumulation of transposons, and/or polyploidy (Ohno, 1970; Cooke *et al*., 1997; Force *et al*., 1999), all of which result in disproportionate methylation (Nagl & Ehrendorfer, 1974; Volpe & Eremenko, 1974; Holliday, 1984; Yoder, Walsh & Bestor, 1997; Matzke & Matzke, 1998; Regev, Lamb & Jablonka, 1998; Colot & Rossignol, 1999; Jones & Takai, 2001; Martienssen & Colot, 2001). This starts and accelerates Muller's ratchet because of the combined effect of longer genome lengths, higher per-locus mutation rates, and lower recombination rates. Therefore, I predict that the ratio of lengths of X and Y (or Z and W) chromosomes in a lineage will be proportional to how derived that lineage is. Derived lineages will have shorter Y or W chromosomes (relative to their respective X or Z chromosomes) compared with ancestral lineages. This prediction is an amalgamation of Muller's ratchet driven by methylation coupled with a generalization of the Steinemanns' retrotransposon trap (Steinemann & Steinemann, 1992, 1998). I generalize their retrotransposon trap by asserting that gene duplications could be of things other than retrotransposons. No other hypothesis predicts that more derived lineages will have shorter sex chromosomes.

This prediction can most easily be tested in amniotic vertebrates. Support comes from snakes, birds, and mammals, which are the only vertebrates with well-differentiated sex chromosomes. Only the most basal lineages of each (boas, ratites and tinamins, monotremes) have poorly differentiated sex chromosomes (Jones & Singh, 1985; Solari, 1994; Graves, 1995a; Pigozzi & Solari, 1999; Pigozzi, 1999). Furthermore, vertebrates with poorly differentiated sex chromosomes have much less banding, indicative of low levels of methylation (Ray-Chaudhuri, Singh & Sharma, 1971; Solari, 1994; Pigozzi & Solari, 1997).

We must also account for genome length in any such analysis because application of the traditional Muller's ratchet without methylation also predicts that derived lineages have shorter Y chromosomes. The speed of Muller's ratchet is proportional to the per-genome point mutation rate, which is itself equal to the average point mutation rate times genome length. Therefore, gene duplications in derived lineages would increase the rate of Muller's ratchet by increasing genome length, even without invoking methylation. To discern whether methylation further increases the speed of Muller's ratchet by also increasing the average point mutation rate and suppressing the recombination rate, one should also estimate genome length at ancestral nodes using phylogenetic comparative methods (Martins & Hansen, 1997; Schluter *et al*., 1997). We could then control for the traditional Muller's ratchet being driven solely by increased genome length.

Vertebrate phylogenies are continually being refined (e.g. Cao *et al*., 2000b; Gissi *et al*., 2000). Data

on vertebrate genome lengths have been compiled in the 'Animal cytogenetics' series (Borganonkar, 1974; Egozcue, 1974; Fregda, 1974; Gustavsson, 1974; Hayman & Martin, 1974; Patton, 1974; Olmo, 1986; Christidis, 1990) and references therein. Phylogenetic comparisons should be especially useful in lineages with a large range of ratios of lengths of their sex chromosomes, such as snakes, iguanids and birds.

Methylation can canalize dioecy

I predict that in lineages in which methylation originally determines the sex of an individual, this methylation also canalizes dioecy. This prediction provides an epigenetic mechanism for inducing canalization.

I predict that lineages with different length sex chromosomes will be dioecious because sex is determined by gene deletions and not just differential methylation. In such individuals, sex changes cannot be induced by altering sex-specific methylation patterns because these patterns disappeared when methylated cytosines mutated to thymines. However, this prediction also arises from other theories.

With equal length sex chromosomes, the prediction that diploid cosexual lineages cannot be descended from dioecious lineages does not arise from any other theory. To test this, correlate dioecy/cosexuality with the proportion of differential chromosomal G-banding, especially where only one sex has heteromorphically banded putative proto-sex chromosomes. Most G-banding seems to be due to heterochromatin, which in turn may be associated with methylation (de Almeida-Toledo *et al*., 1998), so that sex determination is epigenetically controlled (Negrutiu *et al*., 2001). G-banding data (Solari, 1994) and robust phylogenies (Olmo, 1986; Baldauf *et al*., 2000; Cao *et al*., 2000a; Gissi *et al*., 2000) exist, including in lineages for which there is evidence that differential methylation may be the first step in differentiation of sex chromosomes.

It will be valuable to test whether diploid cosexual lineages cannot be descended from dioecious lineages in fish, amphibians, or certain basal reptiles (e.g. boas), and seeing where on the phylogenies these unexpected transitions from dioecy to cosexuality occurred. There are apparently no counterexamples to the prediction that lineages with different length sex chromosomes are dioecious (Solari, 1994).

Although I predict that diploid cosexual lineages cannot be descended from diploid dioecious lineages (Johnston**,** Barnett & Sharpe, 1995), allopolyploidy provides an escape – the only escape – from this canalization of dioecy. Loss of loci from proto-Y or proto-W chromosomes canalizes dioecy in diploid lineages. Allopolyploid progeny, however, can contain chromosomes determining both femaleness and maleness, even though each of their parents could only do one of the two. Flowering plants have much higher incidences of polyploidy than do other seed plants (Otto & Whitton, 2000), hence they should have a higher proportion of cosexual lineages, which would corroborate this prediction. This prediction could be made from several other theories if the diploid dioecious lineage has sex chromosomes of different length. What is unique here is that this prediction cannot be made with other theories if dioecy is solely due to differential methylation of sex-determining genes on one pair of homologous chromosomes, the proto-sex chromosomes, as I believe happens with many plants. A robust test of the prediction that methylation canalizes dioecy is whether allopolyploidy provides an escape from canalization of dioecy.

We could test whether transitions from dioecy to cosexuality always occur along with a polyploid event, particularly in the vast majority of plants in which both sex chromosomes are of equal length. There do not appear to be any counterexamples to the assertion that cosexuality cannot evolve from dioecy unless there is allopolyploidy. This test is vacuous for tetrapods, for which there is no known allopolyploidy (Dowling & Secor, 1997), but should be useful for plants. This test does not suffer from the previously listed problems, but suffers from lack of data on whether chromosomal doubling was due to autopolyploidy or allopolyploidy. However, reversing the logic, this test provides a new means for determining whether allopolyploidy or autopolyploidy was the cause of chromosomal doubling in a lineage.

The escape from canalization test potentially suffers from lack of data on evolutionary transitions from dioecy to cosexuality. It also suffers from the problem that phylogenetic comparative methods all assume no reticulate evolution, a problem that can be sidestepped by using gene phylogenies (which are not reticulate), instead of organism phylogenies.

An important corollary of methylation canalizing dioecy is that temperature-dependent sex determination (TSD) requires homomorphic sex chromosomes. One of the most widely discussed forms of cosexuality in vertebrates is TSD. TSD is a specific form of environmental sex determination (ESD), in which sex determination may be induced by osmotic stress, anomalous levels of oxygen or carbon dioxide, or other environmental factors (Ackerman, 1981; Gutzke & Paukstis, 1983; Spotila**,** Spotila & Kaufer, 1994). TSD and ESD are to be distinguished from genetic sex determination (GSD). I predict that TSD and ESD can only occur in species that have virtually identical sex chromosomes; they must have the same length and have very similar levels of methylation and G-banding. Therefore, small environmental changes can result in alteration of their methylation patterns, thereby determining the sex of each individual. As

soon as sex determination becomes more canalized – either via degenerating length of Y or W chromosomes or via highly disparate differential methylation – then TSD and ESD should not be able to function. Highly heteromorphic sex chromosomes imply that a species must be strictly dioecious, hence have GSD.

A cursory look at reptiles corroborates the prediction that ESD is confined to animals with homomorphic sex chromosomes. The only reptiles with TSD are most turtles, all crocodiles, some geckos, and some agamid lizards (Olmo, 1986; Spotila *et al*., 1994). Each of these seem to have identical or virtually identical sex chromosomes (Olmo, 1986 and references therein). Furthermore, all vertebrates that have Y (or W) chromosomes that are shorter than their X (or Z) chromosomes have GSD (Solari, 1994).

Degeneration of sex chromosomes will be rapid

I predict that degeneration of sex chromosomes will be rapid once loci containing sex-determining methylation patterns incur mutations to (unmethylated) thymine or epimutations to unmethylated cytosine because individuals containing such mutations or epimutations will no longer be recognizable as either sex. Silencing such genes will not help because of their necessary role in sex determination. Degeneration will also be hastened because methylated cytosine has a higher mutation rate. Rapid degeneration of sex chromosomes is contrary to all other models of sex chromosome evolution (Steinemann & Steinemann, 1992; Tucker & Lundrigan, 1995; Rice, 1996; Charlesworth & Charlesworth, 2000).

Searching for evidence of rapid degeneration of sex chromosomes is highly non-trivial, yet it might exist in lineages with incipient sex chromosome formation, such as in some snakes (Jones & Singh, 1985) and fishes (Koehler *et al*., 1995).

I also predict that the rate of degeneration and eventual size of heterogametic sex chromosomes (i.e. Y or W) should be proportional to the total amount of methylation contained on the heterogametic proto-sex chromosome (vice the amount of methylation contained on promoters). Lineages with relatively low overall methylation levels per chromosome will therefore not evolve distinctive sex chromosomes.

Plant and fungal haploid stages will express most of the genes expressed by their diploid stages

Methylation driving Muller's ratchet is the primary cause for evolution of unequal length sex chromosomes in metazoan animals, but not in other sexual diploid eukaryotes, such as plants and fungi. Why? Muller's ratchet simply does not apply to plants and fungi because their haploid genomes are subject to strong selection. The thing that differentiates animals from plants and fungi here is the number of genes expressed in their haploid stages (gametophytes). The haploid stages in two plant species express most of the genes that are expressed by their diploid stages (Willing & Mascarenhas, 1984; Willing, Bashe & Mascarenhas, 1988), thereby nullifying the effects of Muller's ratchet. Similar data do not exist for other plants or fungi. I therefore predict that the haploid stages of most plants and fungi will express most of the genes expressed by their somatic diploid stages (sporophytes). If there are any exceptions to this prediction, they should have sex chromosomes of different lengths.

Microarray analysis, beginning with reverse transcribed mRNA from male and/or female gametophytes and a target treatment with a mixture of sporophyte cells, should readily provide estimates of the percentage of genes expressed in gametophytes (Desprez *et al*., 1998; Schaffer *et al*., 2000). Microarray comparisons of haploid vs. diploid stages have never been done.

Substantial genetic drift implies higher likelihood of cosexuality

I predict that most marine invertebrates with planktonic larvae will have virtually identical sex chromosomes and be hermaphrodites. Many marine invertebrates have much larger effective population sizes than do terrestrial invertebrates, largely due to wide dispersal of larvae on ocean currents. A large effective population size virtually stops Muller's ratchet due to a lack of genetic drift (Nei, 1970; Charlesworth & Charlesworth, 2000). Most marine invertebrates will have virtually identical sex chromosomes and can therefore contain many hermaphroditic species. Terrestrial invertebrates and those marine lineages without ancestral planktonic larvae will have distinctive sex chromosomes and therefore be strictly dioecious. Compounding this predicted effect, most terrestrial invertebrates are evolutionarily derived compared with their marine counterparts. This prediction about relative lengths of sex chromosomes arises from the traditional Muller's ratchet model, except that I have added the prediction about the incidence of hermaphrodites and dioecy.

Testing would require compilation of pairs of closely related species in which one taxon has planktonic larvae and the other has non-planktonic (including terrestrial) larvae and then conducting phylogenetic comparisons of the ratios of sex chromosome lengths with the binary variable of whether the taxa are dioecious or cosexual (Felsenstein, 1985). A cursory glance at marine vs. terrestrial invertebrates seems to indicate a greater incidence of strict dioecy in terres-

trial invertebrates, but this corroboration may be due to the predominance of insects in terrestrial faunas (Ehrlich & Wilson, 1991) and nematodes in deep-sea benthic faunas (Baldwin, Nadler & Hall, 1999). Insects are a highly derived and relatively recently evolved class (Kaesler, 1987), while there are no known nematodes with planktonic larvae (Baldwin *et al*., 1999).

CONCLUSION

Differential methylation of sex chromosomes provides a unifying mechanism by which many patterns of sex determination, sex change, dioecy, and sex chromosome length can be explained. Differential methylation suppresses transcription in a sex-specific manner, causing the origin of two distinct sexes. Methylation suppresses recombination, allowing Muller's ratchet to operate on differentially methylated proto-sex chromosomes. Methylation increases point mutation rates and decreases recombination rates, thereby geometrically accelerating Muller's ratchet. These point mutations (i.e. nucleotide transitions) cause loss of a portion of the methylation that originally determined sex, providing strong selective pressure to excise these mutated and epimutated genes. Duplicated genes, which provide a major evolutionary force, are disproportionally methylated and thereby greatly accelerate Muller's ratchet and the degeneration of heterogametic sex chromosomes.

I have detailed many possible predictions and tests for the hypotheses that methylation was the original determinant of sex and that methylation driving Muller's ratchet was responsible for sex chromosome degeneration. The predictions that derived lineages have smaller ratios of Y to X (or W to Z) chromosome lengths and that methylation canalizes dioecy are readily testable with existing data. The remaining predictions are less readily testable at this juncture because of a current dearth of appropriate data, especially of methylation levels. Nonetheless, there are no technological impediments to conducting these tests, which would then allow robust testing of this theory.

ACKNOWLEDGEMENTS

Thanks to Greg Pollock, Sue Bertram, Roy Osborne, Manfred Laubichler, Alberto Solari, Andrea Case, Jennifer Fewell, Paige Warren, Allen Moore, Todd Katzner, Tom Dowling, Chris Lorson, Carlos Castillo-Chavez, Marty Anderies, Tom Taylor, Michael Berrill, Justen Whittall and two anonymous reviewers for their numerous helpful comments. Thanks to the Arizona State University biology department and graduate college for two summer research fellowships.

REFERENCES

Ackerman RA. 1981. Growth and gas exchange of embryonic sea turtles (*Chelonia, Caretta*). *Copeia* **1981:** 757–765.

de Almeida-Toledo LF, Foresti F, Viegas-Péquignot E, Daniel-Silva MFZ. 2001. XX:XY sex chromosome system with X heterochromatinization: an early stage of sex chromosome differentiation in the Neotropic electric eel *Eigenmannia virescens*. *Cytogenetics and Cell Genetics* **95:** 73–78.

de Almeida-Toledo LF, Viegas-Péquignot E, Coutinho-Barbosa AC, Foresti F, Niveleau A, Toledo-Filho SD. 1998. Localization of 5-methylcytosine in metaphase chromosomes of diploid and triploid pacu fish, *Piaractus mesopotamicus* (Pisces, Characiformes). *Cytogenetics and Cell Genetics* **83:** 21–24.

Arnholdt-Schmitt B, Holzapfel B, Schillinger A, Neumann K-H. 1991. Variable methylation and differential replication of genomic DNA in cultured carrot root explants during growth induction as influenced by hormonal treatments. *Theoretical and Applied Genetics* **82:** 283–288.

Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290:** 972–977.

Baldwin JG, Nadler SA, Hall DH. 1999. Nematodes: pervading the earth and linking all life. In: Raven PH, Williams T, eds. *Nature and human society: the quest for a sustainable world*. Washington DC: National Academy Press, 176–191.

Barrett TE, Savva R, Panayotou G, Barlow T, Brown T, Jiricny J, Pearl LH. 1998. Crystal structure of a G: T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary strand interactions. *Cell* **92:** 117–129.

Borganonkar D. 1974. *Animal cytogenetics: Mammalia II: Placentalia: 4. Insectivora and Chiroptera*. Stuttgart: Gebrüder Borntraeger.

Boyes J, Bird A. 1992. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO Journal* **11:** 327–333.

Brown TC, Jiricny J. 1987. A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. *Cell* **50:** 945–950.

Bull JJ. 1983. *Evolution of sex determining mechanisms*. Menlo Park: Benjamin/Cummings.

Canning CA, Lovell-Badge R. 2002. *Sry* and sex determination: how lazy can it be? *Trends in Genetics* **18:** 111–113.

Cao Y, Fujiwara M, Nikaido M, Okada N, Hasegawa M. 2000a. Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. *Gene* **259:** 149–158.

Cao Y, Sorenson MD, Kumazawa Y, Mindell DP, Hasegawa M. 2000b. Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes. *Gene* **259:** 139–148.

Catcheside DG. 1986. A restriction and modification model for the initiation and control of recombination in *Neurospora*. *Genetical Research* **47:** 157–165.

Charlesworth B. 1978. Model for evolution of Y-chromosomes and dosage compensation. *Proceedings of the National Academy of Sciences, USA* **75:** 5618–5622.

Charlesworth B. 1991. The evolution of sex chromosomes. *Science* **251:** 1030–1033.

Charlesworth B, Charlesworth D. 1978. Model for evolution of dioecy and gynodioecy. *American Naturalist* **112:** 975–997.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences* **355:** 1563–1572.

Christidis L. 1990. *Animal cytogenetics: Aves*. Stuttgart: Gebrüder Borntraeger.

Colot V, Rossignol J-L. 1999. Eukaryotic DNA methylation as an evolutionary device. *Bioessays* **21:** 402–411.

Cooke J, Nowak MA, Boerlijst M, Maynard Smith J. 1997. Evolutionary origins and maintenance of redundant gene expression during metazoan development. *Trends in Genetics* **13:** 360–364.

Darwin CR. 1873. On the males and complemental males of certain cirripedes, and rudimentary structures. *Nature* **8:** 431–432.

Demeulemeester MAC, Van Stallen N, De Proft MP. 1999. Degree of DNA methylation in chicory (*Cichorium intybus* L.): influence of plant age and vernalization. *Plant Science* **142:** 101–108.

Desprez T, Amselem J, Caboche M, Höfte H. 1998. Differential gene expression in *Arabidopsis* monitored using cDNA arrays. *Plant Journal* **14:** 643–652.

Dorazi R, Chesnel A, Dournan C. 1995. Opposite sex determination of gonads in two *Pleurodeles* species may be due to temperature-dependent inactivation of sex chromosomes. *Journal of Heredity* **86:** 28–31.

Dowling TE, Secor CL. 1997. The role of hybrization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics* **28:** 593–619.

Dübendorfer A, Hediger M, Burghardt G, Bopp D. 2002. *Musca domestica*, a window on the evolution of sex-determining mechanisms in insects. *International Journal of Developmental Biology* **46:** 75–79.

Egozcue J. 1974. *Animal cytogenetics: Mammalia II: Placentalia: 5. Primates*. Stuttgart: Gebrüder Borntraeger.

Ehrlich PR, Wilson EO. 1991. Biodiversity studies: science and policy. *Science* **253:** 758–762.

Felsenstein J. 1974. Evolutionary advantage of recombination. *Genetics* **78:** 737–756.

Felsenstein J. 1985. Phylogenies and the comparative method. *American Naturalist* **125:** 1–15.

Felsenstein J. 1988. Sex and the evolution of recombination. In: Michod RE, Levin BR, eds. *The evolution of sex: an examination of current ideas*. Sunderland: Sinauer Associates, 74–86.

Fisher RA. 1931. The evolution of dominance. *Biological Reviews and Biological Proceedings of the Cambridge Philosophical Society* **6:** 345–368.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151:** 1531–1545.

Fregda K. 1974. *Animal cytogenetics: Mammalia II: Placentalia: 2. Carnivora*. Stuttgart: Gebrüder Borntraeger.

Frota-Pessoa O, Aratangy LR. 1968. The degeneration of the Y chromosome. *Revue de Brasileira de Pesquisas Medecina y Biologia* **1:** 241–244.

Futscher BW, Oshiro MM, Wozniak RJ, Holtan N, Hanigan CL, Duan H, Domann FE. 2002. Role for DNA methylation in the control of cell type-specific maspin expression. *Nature Genetics* **31:** 175–179.

Galm O, Rountree MR, Bachman KE, Jair KW, Baylin SB, Herman JG. 2002. Enzymatic regional methylation assay: a novel method to quantify regional CpG methylation density. *Genome Research* **12:** 153–157.

Giancotti P, Grappelli C, Poggesi I, Abatecola M, de Capoa A, Cozzi R, Perticone P. 1995. Persistence of increased levels of ribosomal gene activity in CHO-K1 cells treated in vitro with demethylating agents. *Mutation Research Letters* **348:** 187–192.

Gissi C, Reyes A, Pesole G, Saccone C. 2000. Lineage-specific evolutionary rate in mammalian mtDNA. *Molecular Biology and Evolution* **17:** 1022–1031.

Gitan RS, Shi HD, Chen CM, Yan PS, Huang THM. 2002. Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Research* **12:** 158–164.

Gorelick R. 2003. Transposable elements suppress recombination in all meiotic eukaryotes, including automictic ancient asexuals: a reply to Schön and Martens. *Journal of Natural History* **37:** 903–909.

Gorelick R, Osborne R. 2002. Inducing sex change and organogenesis from tissue culture in the endangered African cycad *Encephalartos woodii* (Cycadales, Zamiaceae). *South African Journal of Science* **98:** 114–117.

Gowher H, Leismann O, Jeltsch A. 2000. DNA of *Drosophila melanogaster* contains 5-methylcytosine. *EMBO Journal* **19:** 6918–6923.

Grant SR. 1999. Genetics of gender dimorphism in higher plants. In: Geber MA, Dawson TE, Delph LF, eds. *Gender and sexual dimorphism in flowering plants*. Berlin: Springer-Verlag, 246–274.

Graves JAM. 1995a. The evolution of mammalian sex chromosomes and the origin of sex determining genes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **350:** 305–312.

Graves JAM. 1995b. The origin and function of the mammalian Y chromosome and Y-borne genes: an evolving understanding. *Bioessays* **17:** 311–320.

Graves JAM. 2002. The rise and fall of *SRY*. *Trends in Genetics* **18:** 259–264.

Griswold MD, Kim J-S. 2001. Site-specific methylation of the promoter alters deoxyribonucleic acid–protein interactions and prevents follicle-stimulating hormone receptor gene transcription. *Biology of Reproduction* **64:** 602–610.

Gustavsson I. 1974. *Animal cytogenetics: Mammalia II: Placentalia: 3. Ungulata*. Stuttgart: Gebrüder Borntraeger.

Gutzke WHN, Paukstis GL. 1983. Influences of hydric envi-

ronment on sexual differentiation on turtles. *Journal of Experimental Zoology* **226:** 467–469.

Haigh J. 1978. The accumulation of deleterious genes in a population: Muller's ratchet. *Theoretical Population Biology* **14:** 251–267.

Hayman DL, Martin PG. 1974. *Animal cytogenetics: Mammalia I: Monotremata and Marsupialia*. Stuttgart: Gebrüder Borntraeger.

Henry I, Forlani S, Vaillant S, Muschler J, Choulika A, Nicolas JF. 1999. LagoZ and LagZ, two genes derived from the LacZ gene to study epigenetics. *Comptes Rendus de l'Academie Des Sciences Serie III – Sciences de la Vie* **322:** 1061–1070.

Holliday R. 1984. The biological significance of meiosis. In: Evans CE, Dickinson HG, eds. *Controlling events in meiosis*. Cambridge: Cambridge University Press, 381–394.

Holliday R. 1988. A possible role for meiotic recombination in germ line reprogramming and maintenance. In: Michod RE, Levin BR, eds. *The evolution of sex: an examination of current ideas*. Sunderland: Sinauer Associates, 45–55.

Holliday R, Ho T. 1991. Gene silencing in mammalian cells by uptake of 5-methyl deoxcytidine 5′ phosphate. *Somatic Cell Molecular Genetics* **17:** 537–542.

Holliday R, Ho T. 1995. Evidence for gene silencing by DNA methylation in normal human diploid fibroblasts. *Somatic Cell Molecular Genetics* **21:** 215–218.

Hsieh P, Meyn S, Camerini-Otero D. 1986. Partial purification and characterization of a recombinase from human cells. *Cell* **44:** 885–894.

Hurst LD, Hamilton WD. 1992. Cytoplasmic fusion and the nature of sexes. *Proceedings of the Royal Society of London Series B-Biological Sciences* **247:** 189–194.

Iannello RC, Gould JA, Young JC, Giudice A, Medcalf R, Kola I. 2000. Methylation-dependent silencing of the testis-specific *Pdha-2* basal promoter occurs through selective targeting of an activating transcription factor/cAMP-responsive element-binding site. *Journal of Biological Chemistry* **275:** 19603–19608.

Iannello RC, Young J, Sumarsono S, Tymms MJ, Dahl HHM, Gould J, Hedger M, Kola I. 1997. Regulation of *Pdha-2* expression is mediated by proximal promoter sequences and CpG methylation. *Molecular and Cellular Biology* **17:** 612–619.

Iguchi-Ariga SMM, Schaffner W. 1989. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes and Development* **3:** 612–619.

Janousek B, Siroky J, Vyskot B. 1996. Epigenetic control of sexual phenotype in a dioecious plant, *Melandium album*. *Molecular and General Genetics* **250:** 483–490.

Johnston CM, Barnett M, Sharpe PT. 1995. The molecular biology of temperature-dependent sex determination. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **350:** 297–303.

Jones KW, Singh L. 1985. Snakes and the evolution of sex chromosomes. *Trends in Genetics* **1:** 55–61.

Jones PA, Rideout WM, Shen JC, Spruck CH, Tsai YC. 1992. Methylation, mutation and cancer. *Bioessays* **14:** 33–36.

Jones PA, Takai D. 2001. The role of DNA methylation in mammalian epigenetics. *Science* **293:** 1068–1070.

Kaesler RL. 1987. Superclass Hexapoda. In: Boardman RS, Cheetham AH, Rowell AJ, eds. *Fossil invertebrates*. Oxford: Blackwell Scientific, 264–269.

Koehler MR, Neuhaus D, Engel W, Schartl M, Schmid M. 1995. Evidence for an unusual ZW/ZW′/ZZ sex-chromosome system in *Scardinius erythrophthalmus* (Pisces, Cyprinidae), as detected by cytogenetic and H-Y antigen analyses. *Cytogenetics and Cell Genetics* **71:** 356–362.

Kuwamura T, Nakashima Y. 1998. New aspects of sex change among reef fishes: recent studies in Japan. *Environmental Biology of Fishes* **52:** 125–135.

Lloyd DG. 1980. Sexual strategies in plants. III. A quantitative method for describing the gender of plants. *New Zealand Journal of Botany* **18:** 103–108.

LoSchiavo F, Pitto L, Giuliano G, Torti G, Nuti-Ronchi V, Marazziti D, Vergara R, Orsdelli S, Terzi M. 1989. DNA methylation of embryogenic carrot cell cultures and its variation as caused by mutation, differentiation, hormones and hypomethylating drugs. *Theoretical and Applied Genetics* **77:** 325–331.

Lyko F, Ramsahoye BH, Jaenisch R. 2000. DNA methylation in *Drosophila melanogaster*. *Nature* **408:** 538–540.

Martienssen RA, Colot V. 2001. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293:** 1070–1074.

Martins EP, Hansen TF. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* **149:** 646–667.

Matzke MA, Matzke AJM. 1998. Polyploidy and transposons. *Trends in Ecology and Evolution* **13:** 241.

McClung CE. 1902. The accessory-chromosome: sex determinant? *Biology Bulletin* **3:** 43–84.

McLachlan JA. 2001. Environmental signalling: what embryos and evolution teach us about endocrine disrupting chemicals. *Endocrine Reviews* **22:** 319–341.

Miniou P, Bourc'his D, Molina Gomes DM, Jeanpierre M, Viegas-Péquignot E. 1997. Undermethylation of *Alu* sequences in ICF syndrome: molecular and in situ analysis. *Cytogenetics and Cell Genetics* **77:** 308–313.

Mittwoch U. 2000. Three thousand years of questioning sex determination. *Cytogenetics and Cell Genetics* **91:** 186–191.

Muller HJ. 1914. A gene for the fourth chromosome of *Drosophila*. *Journal of Experimental Zoology* **17:** 324–336.

Muller HJ. 1964. The relation of recombination to mutation advance. *Mutation Research* **1:** 2–9.

Murphy SK, Jirtle RL. 2000. Imprinted genes as potential genetic and epigenetic toxicologic targets. *Environmental Health Perspectives, Supplements* **108:** 5–11.

Nagl W, Ehrendorfer F. 1974. DNA content, heterochromatin, mitotic index, and growth in perennial and annual *Anthemidea* (Asteraceae). *Plant Systematics and Evolution* **123:** 35–54.

Nakashima Y, Kuwamura T, Yogo Y. 1995. Why be a bothways sex changer? *Ethology* **101:** 301–307.

Negrutiu I, Vyskot B, Barbacar N, Georgiev S, Moneger F. 2001. Dioecious plants: a key to the early events of sex chromosome evolution. *Plant Physiology* **127:** 1418–1424.

Nei M. 1970. Accumulation of nonfunctional genes on sheltered chromosomes. *American Naturalist* **104:** 311–322.

Norberg J, Vihinen M. 2001. Molecular dynamics simulation of the effects of cytosine methylation on structure of oligonucleotides. *Journal of Molecular Structure (Theochem)* **546:** 51–62.

Nyce J. 1991. Gene silencing in mammalian cells by direct incorporation of electroporated 5-methyl-2′ deoxcytidine 5′-phosphate. *Somatic Cell Molecular Genetics* **17:** 543–550.

Ohno S. 1970. *Evolution by gene duplication*. Berlin: Springer-Verlag.

Olmo E. 1986. *Animal cytogenetics: Reptilia*. Stuttgart: Gebrüder Borntraeger.

Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* **34:** 401–437.

Palmer JD, Adams KL, Cho YR, Parkinson CL, Qiu YL, Song KM. 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proceedings of the National Academy of Sciences, USA* **97:** 6960–6966.

Pannell J. 1997. Mixed genetic and environmental sex determination in an androdioecious population of *Mercurialis annua*. *Heredity* **78:** 50–56.

Pardo-Manuel de Villena F, de la Casa-Esperón E, Sapienza C. 2000. Natural selection and the function of genome imprinting: beyond the silenced minority. *Trends in Genetics* **16:** 573–579.

Patton JL. 1974. *Animal cytogenetics: Mammalia II: Placentalia: 1. Rodentia*. Stuttgart: Gebrüder Borntraeger.

Pigozzi LI. 1999. Origin and evolution of the sex chromosomes in birds. *Biocell* **23:** 79–95.

Pigozzi LI, Solari AJ. 1997. Extreme axial equalization and wide distribution of recombination nodules in the primitive ZW pair of *Rhea americana*. *Chromosome Research* **5:** 421–428.

Pigozzi LI, Solari AJ. 1999. The ZW pairs of two paleognath birds from two orders show transitional stages of sex chromosomes. *Chromosome Research* **7:** 541–551.

Proffitt JH, Davie JR, Swinton D, Hattman S. 1984. 5-methylcytosine is not detectable in *Saccharomyces cerevisiae* DNA. *Molecular and Cellular Biology* **4:** 985–988.

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R. 2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences, USA* **97:** 5237–5242.

Rauth S, Song KY, Ayares D, Wallace L, Moore PD, Kucherlapati R. 1986. Transfection and homologous recombination involving single-stranded DNA substrates in mammalian cells and nuclear extracts. *Proceedings of the National Academy of Sciences, USA* **83:** 5587–5591.

Ray-Chaudhuri SP, Singh L, Sharma T. 1971. Evolution of sex chromosomes and formation of W-chromatin in snakes. *Chromosoma* **33:** 239–251.

Regev A, Lamb MJ, Jablonka E. 1998. The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Molecular Biology and Evolution* **15:** 880–891.

Rice WR. 1987a. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex-chromosomes. *Evolution* **41:** 911–914.

Rice WR. 1987b. Genetic hitchhiking and the evolution of reduced genetic activity of the Y chromosome. *Genetics* **116:** 161–167.

Rice WR. 1994. Degeneration of a nonrecombining chromosome. *Science* **263:** 230–232.

Rice WR. 1996. Evolution of the Y sex chromosome in animals. *Bioscience* **46:** 331–343.

Richards EJ, Elgin SCR. 2002. Epigenetic codes for heterochromatin formation and silencing: Rounding up the usual suspects. *Cell* **108:** 489–500.

Riggs AD, Pfeifer GP. 1992. X chromosome inactivation and cell memory. *Trends in Genetics* **8:** 169–174.

Schaffer R, Landgraf J, Pérez-Amador M, Wisman E. 2000. Monitoring genome-wide expression in plants. *Current Opinion in Biotechnology* **11:** 162–167.

Schluter D, Price T, Mooers AO, Ludwig D. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* **51:** 1699–1711.

Schmid M, Haaf T. 1989. Origin and evolution of sex determination in Amphibia: the cytogenetic data. In: Wachtel SS, ed. *Evolutionary mechanisms in sex determination*. Boca Raton: CRC Press, 37–56.

Schmutte C, Yang AS, Beart RW, Jones PA. 1995. Base excision repair of U : G mismatches at a mutational hotspot in the p53 gene is more efficient than base excision repair of T : G mismatches in extracts of human colon tumors. *Cancer Research* **55:** 3742–3746.

Shen J-C, Rideout WM, Jones PA. 1994. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Research* **22:** 972–976.

Singer T, Yordan C, Martienssen R. 2001. Robertson's *Mutator* transposons in *A. thaliana* are regulated by the chromatin-remodeling gene *Decrease in DNA Methylation* (*DDM1*). *Genes and Development* **15:** 591–602.

Sites JW, Archie JW, Cole CJ, Flores-Villela O. 1992. A review of the phylogenetic hypotheses for lizards of the genus *Sceloporus* (Phrynosomatidae): implications for ecological and evolutionary studies. *Bulletin of the American Museum of Natural History* **213:** 1–110.

Solari AJ. 1994. *Sex chromosomes and sex determination in vertebrates*. Boca Raton: CRC Press.

Spotila JR, Spotila LD, Kaufer NF. 1994. Molecular mechanisms of TSD in reptiles: a search for the magic bullet. *Journal of Experimental Zoology* **270:** 117–127.

Stauffer RC, ed. **1975.** *Charles Darwin's natural selection: being the second part of his big book written from 1856 to 1858*. Cambridge: Cambridge University Press.

Steinemann M, Steinemann S. 1992. Degenerating Y chro-

mosome of *Drosophila miranda*: a trap for retrotransposons. *Proceedings of the National Academy of Sciences, USA* **89:** 7591–7595.

Steinemann M, Steinemann S. 1998. Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of *Drosophila Miranda*: a model for sex chromosome evolution. *Genetica* **102/103:** 409–420.

Stevens NM. 1905. Studies in spermatogenesis, with especial reference to the accessory chromosome. Washington DC: Carnegie Institution of Washington, 1–32.

Tate PH, Bird AP. 1993. Effects of DNA on DNA-binding proteins and gene expression. *Current Opinion in Genetics and Development* **3:** 226–231.

Tatra GS, Miranda J, Chinnappa CC, Reid DM. 2000. Effect of light quality and 5-azacytidine on genomic methylation and stem elongation in two ecotypes of *Stellaria longipes*. *Physiologia Plantarum* **109:** 313–321.

Tucker PK, Lundrigan BL. 1995. The nature of gene evolution on the mammalian Y chromosome: lessons from *Sry*. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **350:** 221–227.

Urieli-Shoval S, Gruenbaum Y, Sedat Y, Razin A. 1982. The absence of detectable methylated bases in *Drosophila melanogaster*. *FEBS Letters* **146:** 148–152.

Volpe P, Eremenko T. 1974. Preferential methylation of regulatory genes in HeLa cells. *FEBS Letters* **44:** 121–126.

Vyskot B, Koukalová B, Kovarik A, Sachambul L, Reynolds D, Bezdek M. 1995. Meiotic transmission of a hypomethylated repetitive DNA family in tobacco. *Theoretical and Applied Genetics* **91:** 659–664.

Wachtel SS, Tiersch TR. 1994. The search for the male-determining gene. In: Wachtel SS, ed. *Molecular genetics of sex determination*. San Diego: Academic Press, 1–22.

West SA, Lively CM, Read AF. 1999. A pluralist approach to sex and recombination. *Journal of Evolutionary Biology* **12:** 1003–1012.

Willing RP, Bashe D, Mascarenhas JP. 1988. An analysis of the quantity and diversity of messenger RNAs from pollen and shoots of *Zea mays*. *Theoretical and Applied Genetics* **75:** 751–753.

Willing RP, Mascarenhas JP. 1984. Analysis of the complexity and diversity of mRNAs from pollen and shoots of *Tradescantia*. *Plant Physiology* **75:** 865–868.

Wolffe AP, Matzke MA. 1999. Epigenetics: regulation through repression. *Science* **286:** 481–486.

Woodcock DM, Lawler CB, Linsenmeyer ME, Doherty JP, Warren WD. 1997. Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *Journal of Biological Chemistry* **272:** 7810–7816.

Yang AS, Jones PA, Shibata A. 1996. The mutational burden of 5-methylcytosine. In: Russo VEA, Martienssen R, Riggs AD, eds. *Epigenetic mechanisms in gene regulation*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 77–94.

Yoder JA, Walsh CP, Bestor TH. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics* **13:** 335–340.

Yung R, Ray D, Eisenbraun JK, Deng C, Attwood J, Eisenbraun MD, Johnson K, Miller RA, Hanash S, Richardson B. 2001. Unexpected effects of a heterozygous Dnmt1 null mutation on age-dependent DNA hypomethylation and autoimmunity. *Journals of Gerontology, Series A, Biological Sciences and Medical Sciences* **56:** B268–B276.

APPENDIX C

GENE DUPLICATIONS AND EVOLUTION OF SEX, PART 2,

INDUCING SEX CHANGE

Inducing sex change and organogenesis from tissue culture in the endangered African cycad *Encephalartos woodii* (Cycadales, Zamiaceae)

Root Gorelick[a*] and Roy Osborne[b]

IF INCIPIENT SEX CHROMOSOME DIFFERENTIATION is caused by differential methylation between females and males, then methylating or demethylating cytosine nucleotides may induce sex change. Methylation may also stimulate regeneration of roots and shoots from tissue culture callus and increase genetic variation via greater mutation. We propose using these methods for conserving *Encephalartos woodii*, for which only a single male clone exists, sex change has never been induced, and regeneration from callus tissue has not been accomplished.

The conservation status of the African cycad *Encephalartos woodii* Sander is bleak. All extant plants are derived from a single male clone.[1] Although vegetative propagation is readily accomplished from offsets and more recently from leaf cuttings, genetic diversity is zero.

Osborne[2] proposed two conservation strategies for *E. woodii*: (1) repeated backcrossing of the existing male with females of a closely related species, and (2) chemically or environmentally inducing sex change in tissue-cultured plantlets from the existing male. The first approach resulted in an F_1 generation of *Encephalartos natalensis* × *E. woodii* in the 1980s and recently F_2 seedlings of (*E. natalensis* × *E. woodii*) × *E. woodii* (pers. obs.). We focus on sex change because it is the least well understood of the two approaches and has potentially the greater conservation benefits.

There are potentially six benefits to inducing sex change in *Encephalartos woodii*. First, our proposed method of sex change promotes mutation, thereby increasing genetic variation. Second, there would be increased epigenetic variation due to meiotic recombination.[3] Third, female plants produce megagametophytes and zygotes providing the cells of choice for cycad tissue culture.[4] Fourth, if backcrossing is the only realistic approach to conservation, then it is preferable to use *E. woodii* as the female parent because of maternal inheritance of chloroplast genomes. Fifth, induced sex change may assist in the conservation of other dioecious plants. Sixth, sex change provides fundamental insight into the biology of sex determination. We propose a method for inducing sex change in cycads and regeneration of roots and shoots from callus, including discussion of the pitfalls.

[a]Department of Biology, Arizona State University, Tempe, AZ 85287-1501, U.S.A.

[b]P.O. Box 244, Burpengary, Queensland 4505, Australia.

*Author for correspondence. E-mail: cycad@asu.edu

Theory of sex determination

This article is predicated on the hypothesis that, in both females and males, incipient sex chromosomes arose as a different methylation pattern on one of two homologous autosomes and the altered methylation was of regulatory genetic elements that control sex hormone production. Sexual differentiation in plants is regulated by sex hormones.[5] Cytosine nucleotides are methylated via replacement of the hydrogen at the C-5 position by a methyl group and are highly heritable.[6,7] Methylation causes sex by blocking binding sites for enzymes that mediate transcription of sex hormones.[8–11]

Methylation suppresses transcription in several ways. Heterochromatic proteins bind to methylated cytosines, occupying protein binding sites.[12,13] Methylation alters interactions of histones with promoter regions by stimulating histone deacetylation.[14–16] Bound heterochromatic proteins and histones are called heterochromatin. Transcription is also suppressed because methyl groups are bulky and hydrophobic, thereby changing DNA conformation and blocking binding sites,[17] sometimes converting the normally right-handed DNA helix to left-handed.[18,19]

Dioecious seed plants generally have putative sex chromosomes that are indistinguishable under a light microscope.[20] Close inspection, however, sometimes reveals slight heterochromatin differences,[21,22] which is itself due to differential methylation.[23,24] Therefore, it is distinctly possible that methylation controls sex determination.

Methylation and accompanying heterochromatin can be removed by various factors — such as temperature,[25,26] light,[27] osmotic stress,[28] or hormones[29–31] — resulting in sex change.[32,33] Sex change occurs only in organisms that have (virtually) indistinguishable sex chromosomes, indicating that incipient sex chromosomes are formed by slight differences in methylation. Differential methylation is evolutionarily the first difference between females and males[34] and is the likely cause of reported sex changes in cycads.

Application of theory to sex change in cycads

Although sex-specific differential methylation or heterochromatin has not been examined in cycads, lack of identifiable sex chromosomes[1] and occasional induction of sex change via environmental shocks[35] suggest that sex determination in cycads is due to differential methylation. Definitive evidence for differential demethylation determining sex should be sought in cycads, especially in populations in which some individuals have undergone sex changes, using techniques such as chromomycin staining,[21] high-performance liquid chromatography (HPLC),[36,37] or bisulphide sequencing of promoters of genes that regulate hormone levels.[38–41]

We propose applying demethylating compounds to cycad cells in tissue culture, an approach that has resulted in sex change in at least one angiosperm species.[7] 5-azacytidine, and 5-aza-2′-deoxycytidine demethylate CpG dinucleotides in most eukaryotes,[42–46] while L-ethionine and dihydroxypropyladenine demethylate cytosines in CpNpG trinucleotides of plants (N can be any nucleotide base).[47]

Although demethylating agents have not been used to alter the sex of cycads, we propose that altering methylation of tissue-cultured cells of *Encephalartos woodii* could yield female plants. We suggest trying this procedure first on relatively common species of *Encephalartos*, such as *E. natalensis*. Tissue-cultured angiosperms have survived exposure to 5-azacytidine, although with lower viability and higher mutation rates.[44,48]

Measuring methylation levels using HPLC or bisulphide sequencing before and after application of demethylating compounds would provide a means for

determining appropriate amounts of these chemicals to use in subsequent experiments. The quantity of demethylating compounds could be adjusted so that the resulting diminution of methylation matches that found in cycads of the requisite sex, although we recognize that the 'wrong' methyl groups might get stripped away.[45]

If sex determination in *Encephalartos woodii* is controlled by differential methylation, then there is an *a priori* 50% chance that a male-to-female change can be induced via demethylation. The odds may be lower, however, because nearly all sex changes reported in *Encephalartos* have been from female-to-male (reviewed in ref. 35). If females of *E. woodii* have more heavily methylated promoters of sex determining genes than males, sex change of the extant male clone can be attempted via methylation. Addition of methyl groups to cytosine nucleotides may be possible by treating tissue cultures with 5-methyl deoxcytidine. This approach has been attempted with mammalian cells, where the dose of 5-methyl deoxcytidine was adjusted so that half the cell line survived.[49–51] However, much less experimental work has been done with methylation than with demethylation.

Ideally, the specific nucleotides responsible for regulation of sex hormones and sex should be identified, an assessment made of whether their regulation is controlled by methylation, and then only those specific nucleotides should be selectively methylated and/or demethylated in an attempt to change sex and regenerate roots and shoots from callus. Selective methylation can be attempted using methylated single-stranded oligonucleotides,[52,53] a technique that is not yet fully developed. Until then, attempts should be made by randomly methylating and demethylating the genome, although most such attempts are likely to be unsuccessful.

Regeneration of roots and shoots from tissue culture callus

Even if the sex of *E. woodii* can be changed with demethylating or methylating chemicals, other problems need to be addressed before any protocol can be considered effective, such as growing sex-changed tissue cultures up to full-sized cone-bearing plants, difficulties relating to small effective population size, and evolutionary loss of organelle genomes.

Early attempts to propagate cycads from tissue culture, including *Encephalartos* species, met with limited success (reviewed in ref. 54). Tissue culture of stem, root, and leaf material from several species of *Encephalartos* has given rise to callus.[54,55] Cell cultures from zygotes of two species of *Encephalartos* have been partially successful: shoots were formed, roots were not, and no plants matured.[56] Recent work using tissue from megagametophytes and zygotes of the Mexican cycads *Ceratozamia hildae, C. mexicana, Zamia fischeri, Z. furfuracea, Z. pumila* and particularly *Dioon edule,* has been encouraging, producing plants that can be grown in soil.[4,57,58] Tissue culture of megagametophyte and zygote explants probably leads to more successful regeneration of entire plants because global demethylation and *de novo* methylation may be a prerequisite for proper development, as with most eukaryote embryos.[59–61] An important but as yet unattained objective is to take leaf, stem, or root explants of *Encephalartos* and induce shoot and root formation from calli. This is necessary for E. *woodii,* in which neither megagametophytes nor zygotes presently exist. Another incentive for inducing sex change of the extant male *E. woodii* clone is that megagametophytic and zygotic cells can then be obtained, cultured, and treated with demethylating or methylating agents.

Although regeneration of tissue cultured *E. woodii* has not yet been accomplished, our theoretical framework of sex determination via methylation may provide the method by which differentiation of callus tissue can be induced.[62] Ontogenetic regulation of all plant and animal tissues is largely controlled by demethylation, although many of the methylated genes controlling regulation have not yet been identified. Until now, regeneration of tissue-cultured roots and shoots relied on the application of plant hormones, which work largely by demethylating chromosomal cytosines.[29,31,63] Until we know which genes to demethylate and how to demethylate them selectively, we propose skipping the intermediary of plant hormones and instead applying 5-azacytidine to *E. woodii* callus to induce root and shoot regeneration.[64] We also suggest first applying 5-methyl deoxcytidine to the callus to simulate early embryonic *de novo* methylation.

As with sex change, we expect that regeneration from callus will usually be unsuccessful because of the randomness of methylation and demethylation when applying compounds like 5-azacytidine. Although regulation is largely controlled by methylation of promoters, methylation of downstream regions of genes may also affect regulation.[65] Our best hope is for selective methylation and demethylation or, less elegantly, for many random attempts in the hope that one explant will be successfully regenerated.

The proposed applications of methylating and demethylating agents serve a dual purpose: sex change and regeneration of roots and shoots. The optimal amounts of these chemicals may, however, be different for each function. The targeted loci for methylation may also be different.

Problems following sex change and regeneration

Even if male-to-female sex change can be induced and viable sexual offspring formed in *E. woodii,* there will still be no genetic variability in the species. There will be only epigenetic variation due to different methylation patterns. It is noteworthy that F_1 and F_2 backcrosses of *E. natalensis* × *E. woodii* have been raised.

Methylation-induced mutations, although often deleterious, could be of conservation benefit for *E. woodii* because the surviving mutants increase genetic variability of the population.[61] Methylation creates a tension between increased genetic variation and decreased viability.[66,67] The balance between these two will determine the conservation implications of the proposed methylation/demethylation induced sex change of *E. woodii.* The best possible outcome is that viable sex-changed females will emerge from this protocol carrying a large number of mutations. It is not necessary that all these mutations appear on a single female; the genetic variation of the entire population is critical. A successful protocol for tissue culture of *E. woodii* will produce many sex-changed mutants to be grown into explants with (collectively) as much genetic variation as possible, thereby reducing the probability of extinction.

Cycads have maternally inherited chloroplasts[68] (and mitochondria?). The main disadvantage of backcrossing the existing male *E. woodii* with females of other *Encephalartos* species is that these non-Mendelian organelle genomes from the male parent are irrevocably lost from all backcrosses.[68] A preferred option would be to backcross an induced female *E. woodii* with males of a closely related species so that the *E. woodii* organelle genomes are preserved, providing additional impetus for inducing sex change in *E. woodii.*

Conclusion

If sex is determined by methylation, then attempts should be made to induce sex change of tissue-cultured cycad cells using specific demethylating or methylating agents. A prerequisite is to verify whether differential methylation of specific genes, primarily those producing and regulating hormones, is correlated with sex in cycads. It is desirable also to quantify degree of methylation in cycad clones for which sex change has been reported. Or, better yet, to identify which nucleotide loci contain regulation-controlling methylation.

Establishment of a successful protocol for *in vitro* culture of diploid vegetative or embryonic tissue of *Encephalartos*, particularly *E. woodii*, including organogenesis and the ultimate re-establishment of plantlets in soil, remains a *sine qua non* for this project. As with sex change, the formation of roots and shoots may also be possible by applying methylating and demethylating agents to the callus tissue. This seems especially plausible because plant hormones are known to cause the following: (i) heritable changes in methylation patterns, (ii) sexual differentiation in plants, and (iii) regeneration of roots and shoots from callus.

If sex change is successful with *Encephalartos woodii* and the resulting female plants are sexually viable, then economics will certainly lead to programmes for breeding nascent female clones to the extant male clone. Backcrossing the female clones to males of other *Encephalartos* species should also be encouraged to help overcome the extreme genetic bottleneck for this species unless the methylation or demethylation treatments add substantial genetic variation to the nascent females.

Putting these ideas into practice will require an enormous amount of work and good fortune. Even if attempts to change the sex and regenerate roots and shoots of the male clone of *Encephalartos woodii* are unsuccessful, however, the proposed testing may still provide fundamental insights into the biology of sex determination in seed plants.

We thank Sue Bertram, Bill Martin, and an anonymous reviewer for many helpful comments. Root Gorelick received support for this work via a summer research fellowship from the Graduate College and Biology Department at Arizona State University.

1. Norstog K.J. and Nicholls T.J. (1997). *The Biology of Cycads.* Cornell University Press, Ithaca.
2. Osborne R. (1985). Sex change in cycads — hope for *woodii*. *Encephalartos* **2**, 20–22.
3. Schlichting C.D. and Pigliucci M. (1998). *Phenotypic Evolution: A Reaction Norm Perspective.* *Sinauer Associates, Sunderland, MA.*
4. Chavez V.M. and Litz R.E. (1999). Organogenesis from megagametophyte and zygotic embryo explants of the gymnosperm *Dioon edule* Lindley (Zamiaceae, Cycadales). *Plant Cell Tissue Organ Cult.* **58**, 219–222.
5. Grant S.R. (1999). Genetics of gender dimorphism in higher plants. In *Gender and Sexual Dimorphism in Flowering Plants*, eds M.A. Geber, T.E. Dawson and L.F. Delph, pp. 246–274. Springer-Verlag, Berlin.
6. Holliday R. (1988). A possible role for meiotic recombination in germ line reprogramming and maintenance. In *The Evolution of Sex: An Examination of Current Ideas*, eds R.E. Michod and B.R. Levin, pp. 45–55. Sinauer Associates, Sunderland, MA.
7. Vyskot B., Koukalová B., Kovarik A., Sachambul L., Reynolds D. and Bezdek M. (1995). Meiotic transmission of a hypomethylated repetitive DNA family in tobacco. *Theor. Appl. Genet.* **91**, 659–664.
8. Charlesworth B. and Charlesworth D. (1978). Model for evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997.
9. Griswold M.D. and Kim J-S. (2001). Site-specific methylation of the promoter alters deoxyribonucleic acid-protein interactions and prevents follicle-stimulating hormone receptor gene transcription. *Biol. Reprod.* **64**, 602–610.
10. Iannello R.C., Young J., Sumarsono S., Tymms M.J., Dahl H.H.M. *et al.* (1997). Regulation of *Pdha-2* expression is mediated by proximal promoter sequences and CpG methylation. *Mol. Cell. Biol.* **17**, 612–619.
11. Iannello R.C., Gould J.A., Young J.C., Giudice A., Medcalf R. and Kola I. (2000). Methylation-dependent silencing of the testis-specific *Pdha-2* basal promoter occurs through selective targeting of an activating transcription factor/cAMP-responsive element-binding site. *J. Biol. Chem.* **275**, 19603–19608.
12. Iguchi-Ariga S.M.M. and Schaffner W. (1989). CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev.* **3**, 612–619.
13. Tate P.H. and Bird A.P. (1993). Effects of DNA on DNA-binding proteins and gene expression. *Curr. Opin Genet. Dev.* **3**, 226–231.
14. Davey C., Penning S. and Allan J. (1997). CpG methylation remodels chromatin structure *in vitro*. *J. Mol. Biol.* **267**, 276–288.
15. Jones P.L., Veenstra G.J.C., Wade P.A., Vermaak D., Kass S.U. *et al.* (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Gen.* **19**, 187–191.
16. Nan X.S., Ng H.H., Johnson C.A., Laherty C.D., Turner B.M. *et al.* (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389.
17. Derreumaux S., Chaoui M., Tevanian G. and Fermandjian S. (2001). Impact of CpG methylation on structure, dynamics and solvation of cAMP DNA responsive element. *Nucleic Acids Res.* **29**, 2314–2326.
18. Behe M. and Felsenfeld G. (1981). Effects of methylation on a synthetic polynucleotide: the B-Z transition in poly(dG-m^5dC)·poly(dG-m^5dC). *Proc. Natl. Acad. Sci. USA* **78**, 1619–1623.
19. McKay D.B. and Steitz T.A. (1981). Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature* **290**, 744–749.
20. Chattopadhyay D. and Sharma A.K. (1991). Sex determination in dioecious species of plants. *Feddes Repert.* **102**, 29–75.
21. Siljak-Yakovlev S., Benmalek S., Cerbah M., Coba de la Peña T., Bounaga N. *et al.* (1996). Chromosomal sex determination and heterochromatin structure in date palm. *Sex. Plant Reprod.* **9**, 127–132.
22. Greilhuber J., Ebert I., Lorenz A. and Vyskot B. (2000). Origin of facultative heterochromatin in the endosperm of *Gagea lutea* (Lilliaceae). *Protoplasma* **212**, 217–226.
23. Scarbrough K., Hattman S. and Nur U. (1984). Relationship of DNA methylation level to the presence of heterochromatin in mealybugs. *Mol. Cell. Biol.* **4**, 599–603.
24. Singer T., Yordan C. and Martienssen R. (2001). Robertson's *Mutator* transposons in *A. thaliana* are regulated by the chromatin-remodeling gene *Decrease in DNA methylation* (*DDM1*). *Genes Dev.* **15**, 591–602.
25. Dorazi R., Chesnel A. and Dournan C. (1995). Opposite sex determination of gonads in two *Pleurodeles* species may be due to temperature-dependent inactivation of sex-chromosomes. *J. Hered.* **86**, 28–31.
26. Demeulemeester M.A.C., Van Stallen N. and De Proft M.P. (1999). Degree of DNA methylation in chicory (*Cichorium intybus* L.): influence of plant age and vernalization. *Plant Sci.* **142**, 101–108.
27. Tatra G.S., Miranda J., Chinnappa C.C. and Reid D.M. (2000). Effect of light quality and 5-azacytidine on genomic methylation and stem elongation in two ecotypes of *Stellaria longipes*. *Physiol. Plant.* **109**, 313–321.
28. Kovarik A., Koukalová B., Bezdek M. and Opatrny Z. (1997). Hypermethylation of tobacco heterochromatic loci in response to osmotic stress. *Theor. Appl. Genet.* **95**, 301–306.
29. LoSchiavo F., Pitto L., Giuliano G., Torti G., Nuti-Ronchi V. *et al.* (1989). DNA methylation of embryogenic carrot cell cultures and its variation as caused by mutation, differentiation, hormones and hypomethylating drugs. *Theor. Appl. Genet.* **77**, 325–331.
30. Jost J.P. and Saluz H.P. (1993). Steroid hormone dependent changes in DNA methylation and its significance for the activation or silencing of specific genes. In *DNA Methylation: Molecular Biology and Biological Significance*, eds J.P. Jost and H.P. Saluz, pp. 425–451. Birkhäuser Verlag, Basel.
31. Arnholdt-Schmitt B., Holzapfel B., Schillinger A. and Neumann K-H. (1991). Variable methylation and differential replication of genomic DNA in cultured carrot root explants during growth induction as influenced by hormonal treatments. *Theor. Appl. Genet.* **82**, 283–288.
32. Chandra H.S. (1994). Proposed role of W chromosome inactivation and the absence of dosage compensation in avian sex determination. *Proc. R. Soc. Lond. B* **258**, 79–82.
33. Pannell J. (1997). Mixed genetic and environmental sex determination in an androdioecious population of *Mercurialis annua*. *Heredity* **78**, 50–56.
34. Tucker K.L., Beard C., Dausmann J., Jackson-Grusby L., Laird P.W. *et al.* (1996). Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev.* **10**, 1008–1020.
35. Osborne R. and Gorelick R. (in press). Sex change in cycads. *Palms & Cycads.*
36. Eick D., Fritz H-J. and Doerfler W. (1983). Quantitative determination of 5-methylcytosine in DNA by reverse-phase high-performance liquid chromatography. *Anal. Biochem.* **135**, 165–171.
37. Gehrke C.W. and Kuo K.C. (1984). Quantitative reverse-phase high performance chromatography of major and modified nucleosides in DNA. *J. Chromatogr.* **301**, 199–219.
38. Frommer M., McDonald L.E., Millar D.S., Collis C.M., Watt F. *et al.* (1992). A genomic sequencing protocol which yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1827–1831.
39. Olek A., Oswald J. and Walter J. (1996). A modified and improved method for bisulphite based cytosine methylation analysis. *Nucleic Acids Res.* **24**, 5064–5066.
40. Galm O., Rountree M.R., Bachman K.E., Jair K.W., Baylin S.B. and Herman J.G. (2002). Enzymatic regional methylation assay: a novel method to

quantify regional CpG methylation density. *Genome Res.* **12**, 153–157.
41. Gitan R.S., Shi H.D., Chen C.M., Yan P.S. and Huang T.H.M. (2002). Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res.* **12**, 158–164.
42. Tamame M., Antequera F., Villanueva J.R. and Santos T. (1983). 5-azacytidine induces heritable biochemical and developmental changes in the fungus *Aspergillus niger. J. Gen. Microbiol.* **129**, 2585–2594.
43. Jones P.A. (1985). Altering gene expression with 5-azacytidine. *Cell* **40**, 485–486.
44. Bezdek M., Koukalová B., Brzobohatý B. and Vyskot B. (1991). 5-azacytidine-induced hypomethylation of tobacco HRS60 tandem DNA repeats in tissue-culture. *Planta* **184**, 487–490.
45. Giancotti P., Grappelli C., Poggesi I., Abatecola M., de Capoa A. *et al.* (1995). Persistence of increased levels of ribosomal gene activity in CHO-K1 cells treated in vitro with demethylating agents. *Mutat. Res. Lett.* **348**, 187–192.
46. Albanesi T., Polani S., Cozzi R. and Perticone P. (1999). DNA strand methylation and sister chromatid exchanges in mammalian cells in vitro. *Mutat. Res.* **429**, 239–248.
47. Kovarik A., Koukalová B., Holy A. and Bezdek M. (1994). Sequence-specific hypomethylation of the tobacco genome induced with dihydroxypropyladenine, ethionine and 5-azacytidine. *FEBS Lett.* **353**, 309–311.
48. Fieldes M.A. and Amyot L.M. (1999). Evaluating the potential of using 5-azacytidine as an epimutagen. *Can. J. Bot.* **77**, 1617–1622.
49. Holliday R. and Ho T. (1991). Gene silencing in mammalian cells by uptake of 5-methyl deoxcytidine 5′ phosphate. *Somatic Cell Mol. Genet.* **17**, 537–542.
50. Holliday R. and Ho T. (1995). Evidence for gene silencing by DNA methylation in normal human diploid fibroblasts. *Somatic Cell Mol. Genet.* **21**, 215–218.
51. Nyce J. (1991). Gene silencing in mammalian cells by direct incorporation of electroporated 5-methyl-2′ deoxcytidine 5′-phosphate. *Somatic Cell Mol. Genet.* **17**, 543–550.
52. Holliday R., Ho T. and Paulin R. (1996). Gene silencing in mammalian cells. In *Epigenetic Mechanisms of Gene Regulation*, ed. V.E.A. Russo, R. Martienssen, A.D. Riggs, pp. 47–59. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
53. Schluesener H.J., Seid K., Deininger M. and Schwab J. (2001). Transient *in vivo* activation of rat brain macrophages/microglial cells and astrocytes by immunostimulatory multiple CpG oligonucleotides. *J. Neuroimmunol.* **113**, 89–94.
54. Osborne R. (1990). Micropropagation in cycads. *Mem. NY Bot. Gar.* **57**, 82–88.
55. Koeleman A. and Small J.G.C. (1982). A note on callus formation by stem and root tissue of some *Encephalartos* species. *S. Afr. J. Bot.* **1**, 165–166.
56. Jäger A.K. and van Staden J. (1996). Somatic embryogenesis and organogenesis in *Encephalartos dyerianus* and *E. natalensis. Plant Cell Tissue Organ Cult.* **45**, 99–102.
57. Chavez V.M., Litz R.E. and Norstog K.J. (1992). *In vitro* morphogenesis of *Ceratozamia hildae* and *C. mexicana* from megagametophytes and zygotic embryos. *Plant Cell Tissue Organ Cult.* **30**, 93–98.
58. Chavez V.M., Litz R.E. and Norstog K.J. (1992). Somatic embryogensis and organogenesis in *Zamia fisheri, Z. furfuracea* and *Z. pumila. Plant Cell Tissue Organ Cult.* **30**, 99–105.
59. Holliday R. and Pugh J.E. (1975). DNA modification mechanisms and gene activity during development. *Science* **187**, 226–232.
60. Riggs A.D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* **14**, 9–11.
61. Kaeppler S.M., Kaeppler H.F. and Rhee Y. (2000). Epigenetic aspects of somaclonal variation in plants. *Plant Mol.Biol.* **43**, 179–188.
62. Finnegan E.J., Bretell R.I.S. and Dennis E.S. (1993). The role of DNA methylation in the regulation of plant gene expression. In *DNA Methylation: Molecular Biology and Biological Significance*, eds J.P. Jost and H.P. Saluz, pp. 218–261. Birkhäuser Verlag, Basel.
63. McLachlan J.A., Burow M., Chiang T.C. and Li S.F. (2001). Gene imprinting in developmental toxicology: a possible interface between physiology and pathology. *Toxicol. Lett.* **120**, 161–164.
64. Amasino R.M., Powell A.L.T. and Gordon M.P. (1984). Changes in T-DNA methylation and expression are associated with phenotypic variation and plant regeneration in a crown gall tumor line. *Mol. Gen. Genet.* **197**, 437–446.
65. Jones P.A. (1999). The DNA methylation paradox. *Trends Genet.* **15**, 34–37.
66. Parra R., Pastor M.T., Perez-Paya E. and Amo-Marco J.B. (2001). Effect of in vitro shoot multiplication and somatic embryogenesis on 5-methylcytosine content in DNA of *Myrtus communis* L. *Plant Growth Regul.* **33**, 131–136.
67. Phillips R.L., Kaeppler S.M. and Olhoft P. (1994). Genetic instability of plant-tissue cultures: breakdown of normal controls. *Proc. Natl. Acad. Sci. USA.* **91**, 5222–5226.
68. Cafasso D., Cozzolino S., Caputo P. and De Luca P. (2001). Maternal inheritance of plastids in *Encephalartos* Lehm. (Zamiaceae, Cycadales). *Genome* **44**, 239–241.

APPENDIX D

GENE DUPLICATIONS AND EVOLUTION OF SEX, PART 3,

AUTOMICTIC ANCIENT ASEXUAL ANIMALS

JOURNAL OF NATURAL HISTORY, 2003, **37,** 903–909

Transposable elements suppress recombination in all meiotic eukaryotes, including automictic ancient asexuals: a reply to Schön and Martens

ROOT GORELICK

Department of Biology, Arizona State University, Tempe, AZ 85287-1501, USA; e-mail: cycad@asu.edu

(*Accepted 31 May 2002*)

I discuss three problems with the argument that ancient asexual lineages lack meiosis and recombination resulting in a diminution of (potentially disadvantageous) transposable elements and other forms of repetitive DNA. First, all ancient asexuals are probably automictic, hence could retain meiotic recombination in the guise of gene conversion. Second, transposable elements alter recombination rates, not vice versa. Third, increasing the number of transposable elements and other forms of repetitive DNA reduces meiotic recombination rate. Ancient asexuals lack those transposable elements that are transmitted via outcrossing, and this has nothing to do with meiotic recombination.

KEYWORDS: Methylation, 5-methylcytosine, heterochromatin, parthenogenetic, bdelloid, darwinulid, *Darwinula*, *Artemia*, automixis.

Introduction

In their stimulating paper on the role of transposable elements in lineages of ancient asexual metazoans, Schön and Martens (2002) argue that lack of sexuality implies lack of meiosis and recombination, which in turn causes a diminution in the number of costly transposable elements and other forms of repetitive DNA. Although their paper is remarkably informative and provides a nice review of ancient asexual metazoans and the possible role of transposable elements and other forms of highly repetitive DNA in the continued existence and asexuality of these lineages, I wish to point out three crucial problems with their argument. First, there is growing evidence that all ancient asexuals are automictic, hence could retain meiotic recombination in the form of gene conversion even if chromosomes are no longer homologous. Second, transposable elements alter recombination rates, not vice versa. Third, a preponderance of transposable elements and other forms of repetitive DNA results in a reduced meiotic recombination rate, i.e. transposable element number and recombination rate are negatively correlated. I elaborate on all three of these problems below and, in so doing, introduce all known extant ancient asexual lineages and the possible evolutionary consequences of their peculiar mode of reproduction.

Journal of Natural History
ISSN 0022-2933 print/ISSN 1464-5262 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0022293021000007705

Ancient asexual lineages are those that have survived without any outcrossing for tens of millions of years. There are probably only three such ancient asexual animal lineages: bdelloid rotifers, darwinulid ostracods, and one or a few lineages of parthenogenetic brine shrimp (Judson and Normark, 1996). Lack of outcrossing causes species concepts to be a bit muddled for asexual lineages. Nonetheless, these are well-defined lineages with facultatively or obligately outcrossing sister taxa.

Ancient asexuals are automictic

The term ancient asexual conveys a misleading connotation. All ancient asexuals probably underwent and still undergo meiosis, albeit automixis, either occasionally or obligately each generation. Automixis refers to production of functional gametes (eggs) via normal meiosis, but pairs of gametes from the same individual then fuse to form a zygote. Thus, antithetically, ancient asexuals are meiotic, but essentially self-fertilizing from pairs of eggs. They never outcross.

There are two forms of meiotic recombination: crossing over and gene conversion. In sexual lineages both forms of recombination are roughly equally prevalent (Carpenter, 1987; Carpenter, 1994). Lineages that have evolved without any form of meiosis for tens of millions of years are predicted to have widely divergent homologues (Welch and Meselson, 2000), to the point where each pair of chromosomes can scarcely be termed homologous. However, this prediction also applies to any lineage that has been without crossing over recombination for tens of millions of years, even if there is obligate or facultative automixis. Unlike crossing over, gene conversion can still occur on non-homologous pairs of chromosomes (Carpenter, 1994). Gene conversion is mediated by early recombination nodules, which can bind to non-homologous segments of DNA. Crossing over is mediated by late recombination nodules, which can only bind to homologous DNA strands (Carpenter, 1994). Therefore, gene conversion is far more likely than crossing over in ancient asexuals.

Parthenogenetic *Artemia* brine shrimp have been asexual for approximately 30 million years (Browne, 1992) and are automictic (Barigozzi, 1974, citing Stefani, 1960). Curiously, polyploid lineages of parthenogenetic *Artemia* are apomictic (Browne, 1992).

Freshwater ostracods in the family Darwinulidae have been asexual for at least the past 70 million years (Butlin and Griffiths, 1993). Darwinulid ostracods probably have occasional facultative automixis (Butlin *et al.*, 1998). This conclusion is based on lack of divergence of ITS1 (internal transcribed spacer 1) sequences, in which there is virtually no allelic divergence. Although this inference is based on a single locus, it runs completely contrary to expectations from any model of ancient obligate apomixes (e.g. Welch and Meselson, 2000). Butlin *et al.* (1998) suggest that darwinulid ostracods' lack of allelic divergence could be due to somatic recombination, gene conversion or automixis. Mitotic recombination is possible, but has never been regarded as an important evolutionary force in animals. The other two possibilities, gene conversion and automixis, are essentially synonymous. Gene conversion requires some form of meiosis, whilst automixis seems to be inevitable if there is any form of meiosis. Thus, it is highly likely, albeit not yet proven, that darwinulid ostracods are automictic.

Rotifers in the family Bdelloidea have been asexual for at least the past 35 million years (Welch and Meselson, 2000). Although there is no definitive report of automixis in bdelloid rotifers, wholesale reset of their epigenetic inheritance systems each generation (Ricci *et al.*, 1999) is completely commensurate with existence of

meiosis and syngamy. In eukaryotes, wholesale epigenetic resets are only associated with gamete formation and following zygote formation, i.e. meiosis and syngamy (Jue *et al.*, 1995; Lei *et al.*, 1996; Shemer *et al.*, 1996; Bestor, 1998; Mertineit *et al.*, 1998). Bdelloid rotifers have experienced huge divergences between formerly homologous alleles, indicating that they never undergo outcrossing meiosis with crossing over recombination. However, despite an early report that bdelloid rotifers are apomictic (Pagani *et al.*, 1993), epigenetic reset data strongly indicate that bdelloids are automictic. Discovery of gene conversion would corroborate this inference of automixis in bdelloids.

Transposable elements suppress crossing over

Schön and Martens (2002) are not alone in asserting that lack of sexuality and recombination results in a diminution in the number of (potentially costly) transposable elements and other forms of repetitive DNA. Virtually all literature on this subject ascribes causality from sexuality and recombination to numbers of transposable elements (e.g. Charlesworth *et al.*, 1994; Charlesworth and Charlesworth, 1995), based largely on population genetic arguments. However, there is a preponderance of molecular evidence showing that the causality is, in fact, reversed. Numbers of transposable elements and other forms of repetitive DNA affect rates of meiotic recombination, and not vice versa.

Transposable elements and repetitive DNA affect recombination because they are disproportionally heavily methylated compared with other portions of the genome. Duplicated, translocated, or inverted portions of the genome are highly methylated (Volpe and Eremenko, 1974; Holliday, 1984; Yoder *et al.*, 1997; Matzke and Matzke, 1998; Regev *et al.*, 1998; Jones and Takai, 2001). This is a consequence of methylation having probably evolved as a defence against parasitic insertion of foreign DNA into the genome (Bestor, 1990; Yoder *et al.*, 1997). In addition to suppressing transcription for genomic defence, methylation also suppresses recombination. Recombination and transcription are suppressed by methylation blocking binding sites for proteins that mediate recombination and transcription, respectively (Catcheside, 1986; Hsieh *et al.*, 1986; Rauth *et al.*, 1986).

Methylation blocks protein binding sites in several inter-related ways. Methyl groups are bulky and hydrophobic and can directly block protein binding sites. Heterochromatic proteins bind to methylated cytosines, which then occupy protein binding sites, an effect that is enhanced by conformational changes in methylated DNA. Methylation restricts the conformational space of the major groove of DNA because methyl groups are bulky (Derreumaux *et al.*, 2001). Protein binding is inhibited by methylation-induced conformational changes, sometimes even to the point of converting the normal right-handed DNA helix to a left-handed helix (Behe and Felsenfeld, 1981; Behe *et al.*, 1981; McKay and Steitz, 1981). Methylation alters the folding of DNA because methyl groups are hydrophobic (Derreumaux *et al.*, 2001) and affects base stacking ability (Norberg and Vihinen, 2001). Methylation alters interactions of histones with promoter regions by stimulating histone deacetylase activity (Davey *et al.*, 1997; Jones *et al.*, 1998; Nan *et al.*, 1998; Henry *et al.*, 1999) and the interactions of DNA with itself (Mayer-Jung *et al.*, 1997), such as supercoiling, thereby affecting protein binding, transcription, and recombination.

Data on methylation suppressing recombination appear to only exist for crossing over (Holliday, 1984; Holliday, 1988; Colot and Rossignol, 1999), and not gene conversion. Crossing over recombination is suppressed in portions of the genome

with transposable elements and with copious highly repetitive DNA, and these portions of the genome have relatively large amounts of heterochromatin (John and Miklos, 1979), which provides an implicit indication of methylation. Lack of data regarding suppression of gene conversion is understandable because there is much more interest in crossing over. Thus, we must rely on a theoretical argument. Proteins (two different forms of recombination nodules) mediate both forms of recombination, and their actions are probably suppressed by methylation. However, gene conversion lacks the chiasmata formed in crossing over (Carpenter, 1994), hence gene conversion requires a smaller set of proteins. Therefore, methylation probably suppresses both gene conversion and crossing over recombination, but possibly does not suppress gene conversion to as great a degree as it suppresses crossing over.

Portions of the genome with a preponderance of transposable elements and repetitive DNA are more highly methylated than the rest of the genome. Therefore, recombination rate is negatively correlated with number of transposable elements and other forms of repetitive DNA because methylation suppresses recombination.

Contrary to the above theory and data of a negative correlation between transposable elements and recombination rates, Schön and Martens (2002: 384) state that 'Density of selfish DNA correlates positively with rate of recombination'. They readily admit that their assertion is based on a subset of transposable elements. It seems, however, that the most likely explanation for their putative positive correlation is that several classes of transposable elements are only transmitted sexually, hence are absent from ancient asexuals (Hickey, 1982; Arkhipova and Meselson, 2000). Schön and Martens' (2002) putative positive correlation between recombination rate and numbers of transposable elements appears to be based on the fact that ancient asexual individuals do not outcross, hence do not transmit certain classes of so-called selfish DNA. A broader look at all transposable elements (including those that are transmitted horizontally by pathogens) and highly repetitive DNA sequences (such as microsatellites) would undoubtedly reveal this to be a negative correlation, commensurate with theory.

Schön and Martens' (2002) paper contains a short discussion of Duret *et al.*'s (2000) nematode work showing that certain transposons are located preferentially in regions of high recombination, focusing on correlations. Instead, I focus on the causality underlying Duret *et al.*'s results. There is persuasive evidence that mobile genetic elements, such as transposons and repeats, are preferentially inserted into highly hypomethylated portions of the genome (Waugh O'Neill *et al.*, 1998; Ehrlich, 2000; Tuck-Muller *et al.*, 2000; Waugh O'Neill *et al.*, 2001). However, once these nascent genetic elements are inserted, they then become highly hypermethylated (Volpe and Eremenko, 1974; Deumling, 1981; Matzke and Matzke, 1998; Regev *et al.*, 1998; Jones and Takai, 2001). If the newly inserted portions of the genome form a relatively small portion of the genomic region, then that region will still be comparatively under methylated, and therefore will experience high recombination rates. Note that, with this argument, it does not matter whether the DNA was inserted into coding or non-coding portions of the genome. Thus, the results of Duret *et al.* (2000) are entirely consistent with the claim that, *ceteris paribus*, transposons are disproportionally highly methylated and methylation suppresses recombination.

Evolutionary implications

For ancient asexual lineages, what are the consequences of lower numbers of transposable elements due to lack of sexual outcrossing and the consequent increase

in recombination rates? Transposable elements are generally considered deleterious. However, they can provide sources for heritable variation, such as somoclonal variation in plants (Kaeppler *et al.*, 2000), and hence possibly allow individuals to survive environmental exigencies better (Griffiths and Butlin, 1995). It is not obvious how this balance between deleterious and beneficial effects of the dearth of sexually transmitted transposable elements affects the evolutionary trajectory of ancient asexual lineages.

Automictic ancient asexual lineages retain the ability to reset epigenetic inheritance systems each generation and to engage in gene conversion, an oft-forgotten form of meiotic recombination. These are two potent evolutionary forces, both of which probably allow these lineages to survive without any outcrossing or crossing over recombination. Furthermore, the potential increased recombination rate in ancient asexuals due to decreased methylation may increase the probability of gene conversion. Compared with apomixes, automixis clearly confers an evolutionary advantage.

Are ancient asexuals less burdened? Perhaps Schön and Martens (2002) are correct that ancient asexuals have escaped the costs of many classes of transposable elements and repetitive DNA. Schön and Martens provide some fascinating and persuasive arguments. I only wish to temper their arguments by tinkering with the details. In particular, I wish to highlight that ancient asexuals probably do undergo meiosis and syngamy—albeit automixis—and their lack of certain classes of transposable elements is due to lack of vertical transmission via outcrossing, and not due to suppressed recombination.

References

Arkhipova, I. and Meselson, M., 2000, Transposable elements in sexual and ancient asexual taxa, *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 14473–14477.

Barigozzi, C., 1974, *Artemia*: a survey of its significance in genetics problems, *Evolutionary Biology*, **7**, 221–252.

Behe, M. and Felsenfeld, G., 1981, Effects of methylation on a synthetic polynucleotide: the B–Z transition in poly(dG-m^{5}dC)·poly(dG-m^{5}dC), *Proceedings of the National Academy of Sciences of the United States of America*, **78**, 1619–1623.

Behe, M., Zimmerman, S. and Felsenfeld, G., 1981, Changes in the helical repeat of poly(dG-m^{5}dC)·poly(dG-m^{5}dC) and poly(dG-dC)·poly(dG-dC) associated with the B–Z transition, *Nature*, **293**, 233–235.

Bestor, T. H., 1990, DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes, *Philosophical Transactions of the Royal Society of London Series B*, **326**, 179–187.

Bestor, T. H., 1998, Cytosine methylation and the unequal developmental potentials of the oocyte and sperm genomes, *American Journal of Human Genetics*, **62**, 1269–1273.

Browne, R. A., 1992, Population genetics and ecology of *Artemia*: insights into parthogenetic reproduction, *Trends in Ecology and Evolution*, **7**, 232–237.

Butlin, R. K. and Griffiths, H. I., 1993, Aging without sex, *Nature*, **364,** 680.

Butlin, R., Schön, I. and Martens, K., 1998, Asexual reproduction in nonmarine ostracods, *Heredity*, **81**, 473–480.

Carpenter, A. T. C., 1987, Gene conversion, recombination nodules, and the initiation of meiotic synapsis, *BioEssays*, **6**, 232–236.

Carpenter, A. T. C., 1994, The recombination nodule story: seeing what you are looking at, *BioEssays*, **16**, 69–74.

Catcheside, D. G., 1986, A restriction and modification model for the initiation and control of recombination in *Neurospora*, *Genetical Research*, **47**, 157–165.

Charlesworth, D. and Charlesworth, B., 1995, Transposable elements in inbreeding and outbreeding populations, *Genetics*, **140**, 415–417.

Charlesworth, B., Sniegowski, P. and Stephan, W., 1994, The evolutionary dynamics of repetitive DNA in eukaryotes, *Nature*, **371**, 215–220.

Colot, V. and Rossignol, J.-L., 1999, Eukaryotic DNA methylation as an evolutionary device, *BioEssays*, **21**, 402–411.

Davey, C., Pennings, S. and Allan, J., 1997, CpG methylation remodels chromatin structure in vitro, *Journal of Molecular Biology*, **267**, 276–288.

Derreumaux, S., Chaoui, M., Tevanian, G. and Fermandjian, S., 2001, Impact of CpG methylation on structure, dynamics and solvation of cAMP DNA responsive element, *Nucleic Acids Research*, **29**, 2314–2326.

Deumling, B., 1981, Sequence arrangement of a highly methylated satellite DNA of a plant, *Scilla*: a tandemly repeated inverted repeat, *Proceedings of the National Academy of Sciences of the United States of America*, **78**, 338–342.

Duret, L., Marais, G. and Biémont, C., 2000, Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*, *Genetics*, **156**, 1661–1669.

Ehrlich, M., 2000, DNA methylation: normal development, inherited diseases, and cancer, *Journal of Clinical Ligand Assay*, **23**, 144–146.

Griffiths, H. I. and Butlin, R. K., 1995, A timescale for sex versus parthenogenesis: evidence from subfossil ostracods, *Proceedings of the Royal Society of London Series B*, **260**, 65–71.

Henry, I., Forlani, S., Vaillant, S., Muschler, J., Choulika, A. and Nicolas, J. F., 1999, LagoZ and LagZ, two genes derived from the LacZ gene to study epigenetics, *Comptes Rendus De L'Academie Des Sciences Serie III*, **322**, 1061–1070.

Hickey, D. A., 1982, Selfish DNA: a sexually-transmitted nuclear parasite, *Genetics*, **101**, 519–531.

Holliday, R., 1984, The biological significance of meiosis, in Evans, C. E. and Dickinson, H. G. (eds) *Controlling events in meiosis* (*Society for Experimental Biology Symposia, no. 38*) (Cambridge: Cambridge University Press), pp. 381–394.

Holliday, R., 1988, A possible role for meiotic recombination in germ line reprogramming and maintenance, in Michod, R. E. and Levin, B. R. (eds) *The evolution of sex: an examination of current ideas* (Sunderland: Sinauer Associates), pp. 45–55.

Hsieh, P., Meyn, S. and Camerini-Otero, D., 1986, Partial purification and characterization of a recombinase from human cells, *Cell*, **44**, 885–894.

John, B. and Miklos, G. L. G., 1979, Funtional aspects of satellite DNA and heterochromatin, *International Review of Cytology*, **58**, 1–114.

Jones, P. A. and Takai, D., 2001, The role of DNA methylation in mammalian epigenetics, *Science*, **293**, 1068–1070.

Jones, P. L., Veenstra, G. J. C., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J. and Wolffe, A. P., 1998, Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription, *Nature Genetics*, **19**, 187–191.

Judson, O. P. and Normark, B. B., 1996, Ancient asexual scandals, *Trends in Ecology and Evolution*, **11**, A41–A46.

Jue, K., Bestor, T. H. and Trasler, J. M., 1995, Regulated synthesis and localization of DNA methyltransferase during spermatogenesis, *Biology of Reproduction*, **53**, 561–569.

Kaeppler, S. M., Kaeppler, H. F. and Rhee, Y., 2000, Epigenetic aspects of somaclonal variation in plants, *Plant Molecular Biology*, **43**, 179–188.

Lei, H., Oh, S. P., Okano, M., Juttermann, R., Goss, K. A., Jaenisch, R. and Li, E., 1996, De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells, *Development*, **122**, 3195–3205.

Matzke, M. A. and Matzke, A. J. M., 1998, Polyploidy and transposons, *Trends in Ecology and Evolution*, **13**, 241.

Mayer-Jung, C., Moras, D. and Timsit, Y., 1997, Effect of cytosine methylation on DNA–DNA recognition at CpG steps, *Journal of Molecular Biology*, **270**, 328–335.

McKay, D. B. and Steitz, T. A., 1981, Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA, *Nature*, **290**, 744–749.

Mertineit, C., Yoder, J. A., Taketo, T., Laird, D. W., Trasler, J. M. and Bestor, T. H., 1998, Sex-specific exons control DNA methyltransferase in mammalian germ cells, *Development*, **125**, 889–897.

Nan, X. S., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N. and Bird, A. P., 1998, Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex, *Nature*, **393**, 386–389.

Norberg, J. and Vihinen, M., 2001, Molecular dynamics simulation of the effects of cytosine methylation on structure of oligonucleotides, *Journal of Molecular Structure* (*Theochem*), **546**, 51–62.

Pagani, M., Ricci, C. and Redi, C. A., 1993, Oogenesis in *Macrotrachela quadrornifera* (Rotifera, Bdelloidea). 1. Germarium eutely, karyotype and DNA content, *Hydrobiologia*, **255**, 225–230.

Rauth, S., Song, K. Y., Ayares, D., Wallace, L., Moore, P. D. and Kucherlapati, R., 1986, Transfection and homologous recombination involving single-stranded DNA substrates in mammalian cells and nuclear extracts, *Proceedings of the National Academy of Sciences of the United States of America*, **83**, 5587–5591.

Regev, A., Lamb, M. J. and Jablonka, E., 1998, The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Molecular Biology and Evolution*, **15**, 880–891.

Ricci, C., Santo, N., Radaelli, E. and Bolzern, A. M., 1999, Epigenetic inheritance systems in bdelloid rotifers. I. Maternal-age-related biochemical effects, *Italian Journal of Zoology*, **66**, 333–339.

Schön, I. and Martens, K., 2002, Are ancient asexuals less burdened? Selfish DNA, transposons and reproductive mode, *Journal of Natural History*, **36**, 379–390.

Shemer, R., Birger, Y., Dean, W. L., Reik, W., Riggs, A. D. and Razin, A., 1996, Dynamic methylation adjustment and counting as part of imprinting mechanisms, *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 6371–6376.

Stefani, R., 1960, L'*Artemia salina* parthenogenitica a Cagliari, *Rivista Di Biologia*, **52**, 463–491.

Tuck-Muller, C. M., Narayan, A., Tsien, F., Smeets, D. F. C. M., Sawyer, J., Fiala, E. S., Sohn, O. S. and Ehrlich, M., 2000, DNA hypomethylation and unusual chromosome instability in cell lines from ICF syndrome patients, *Cytogenetics and Cell Genetics*, **89**, 121–128.

Volpe, P. and Eremenko, T., 1974, Preferential methylation of regulatory genes in HeLa cells, *FEBS Letters*, **44**, 121–126.

Waugh O'Neill, R. J., O'Neill, M. J. and Graves, J. A. M., 1998, Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid, *Nature*, **393**, 68–72.

Waugh O'Neill, R. J., Eldridge, M. D. B. and Graves, J. A. M., 2001, Chromosome heterozygosity and de novo chromosome rearrangements in mammalian interspecies hybrids, *Mammalian Genome*, **12**, 256–259.

Welch, D. M. and Meselson, M. S., 2000, Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange, *Science*, **288**, 1211–1215.

Yoder, J. A., Walsh, C. P. and Bestor, T. H., 1997, Cytosine methylation and the ecology of intragenomic parasites, *Trends in Genetics*, **13**, 335–340.

APPENDIX E

DECOMPOSING MULTI-LOCUS LINKAGE DISEQUILIBRIUM

Decomposing Multilocus Linkage Disequilibrium

Root Gorelick[1] and Manfred D. Laubichler

Department of Biology, Arizona State University, Tempe, Arizona 85287-1501

Manuscript received November 26, 2003
Accepted for publication December 18, 2003

ABSTRACT

We present a mathematically precise formulation of total linkage disequilibrium between multiple loci as the deviation from probabilistic independence and provide explicit formulas for all higher-order terms of linkage disequilibrium, thereby combining J. Dausset *et al.*'s 1978 definition of linkage disequilibrium with H. Geiringer's 1944 approach. We recursively decompose higher-order linkage disequilibrium terms into lower-order ones. Our greatest simplification comes from defining linkage disequilibrium at a single locus as allele frequency at that locus. At each level, decomposition of linkage disequilibrium is mathematically equivalent to number theoretic compositions of positive integers; *i.e.*, we have converted a genetic decomposition into a mathematical decomposition.

A precise measurement of linkage disequilibrium is required for studying virtually any phenomenon in multilocus population genetics. This is especially true for explicit multilocus models that investigate the contributions of physiological epistasis to additive genetic variance (Cheverud and Routman 1995; Wagner *et al.* 1998; Wagner and Laubichler 2000). Linkage disequilibrium is usually defined as the deviation from probabilistic independence between alleles at two different loci. This deviation from independence can have different causes, such as a lack of independent segregation or recombination, or any number of other evolutionary forces. The presence of linkage disequilibrium (gametic disequilibrium) is thus an indication that either stochastic (*e.g.*, drift) or deterministic (*e.g.*, selection, gene flow) evolutionary forces have been acting on a population (Hedrick 2000; Ardlie *et al.* 2002).

The classical definition of linkage disequilibrium, D, follows the probability theory definition of deviation from independence. Independence of two events, B and C, means that $\Pr(BC) = \Pr(B) \cdot \Pr(C)$, where Pr is probability and BC is the joint distribution of B and C, so that the deviation from independence is measured as $D = \Pr(BC) - \Pr(B) \cdot \Pr(C)$. Changing notation slightly to let $A_{k(i)}$ designate the kth allele at the ith locus gives the linkage disequilibrium between the alleles at two loci, D_2, as $D_2 = \Pr(A_{k(1)}A_{k(2)}) - \Pr(A_{k(1)}) \cdot \Pr(A_{k(2)})$, where Pr represents probability and $A_{k(1)}A_{k(2)}$ represents the joint occurrence of $A_{k(1)}$ and $A_{k(2)}$ in a single haploid gamete. In most modern interpretations of probability theory, the primitive concept of "probability" is interpreted as a relative frequency; therefore, $\Pr(A_{k(1)})$ is the same as the frequency of allele k at locus 1.

[1]*Corresponding author:* Department of Biology, Arizona State University, P.O. Box 871501, Tempe, AZ 85287-1501.
E-mail: cycad@asu.edu

The quintessential examples of linkage disequilibrium are coadapted gene complexes, in which several loci are tightly linked because they provide a large selective advantage if they occur together. In these cases, linkage disequilibrium is maintained by selection. Although coadapted gene complexes are implicit in Wright's shifting-balance hypothesis (Wright 1931), have been used to explain outbreeding depression (Dobzhansky 1948; Lynch 1991), and are frequently cited as evolutionary hypotheses (Palopoli and Wu 1996; Rawson and Burton 2002), the linkage disequilibrium of these purported coadapted gene complexes is almost never quantified. This is particularly surprising given the well-cited article by Geiringer (1944), in which she provides most of the algorithm for computing higher-order linkage disequilibrium coefficients. In this article, we complete and simplify Geiringer's formulation and then show how the sums of products of those coefficients equal the definition of (total) linkage disequilibrium as the deviation from probabilistic independence given by Dausset *et al.* (1978).

Methodologically, we follow Geiringer's lead and decompose higher-order linkage disequilibrium into lower-order linkage disequilibrium terms. In other words, we take a top-down approach to defining multilocus linkage disequilibrium, rather than the bottom-up approach followed by virtually everyone since Geiringer (1944). Lewontin (1974) is typical of the bottom-up approach. There are very few other top-down decomposition approaches such as Bulmer's (1980) decomposition of multilocus epistasis or Wagner and Laubichler's (2000) character decomposition approach in population genetics.

In this article, we first define linkage disequilibrium at a single locus as the allele frequency at this locus, which greatly simplifies notation. Second, we extend the definition of linkage disequilibrium to multiple loci

by invoking compositions of positive integers. Our decomposition of multilocus linkage disequilibrium is entirely consistent with the standard definitions for two loci, as well as its previous extensions to three, four, and six loci (Geiringer 1944; Bennett 1954; Hastings 1984). Third, we show how this definition is entirely consistent with the notion of linkage disequilibrium as the deviation from probabilistic independence.

Decomposition of Multilocus Linkage Disequilibrium

Define the one-locus coefficient of linkage disequilibrium, D_1, as $D_1(A_{k(i)}) = \Pr(A_{k(i)})$. This definition may appear paradoxical, but it dramatically simplifies notation for the decomposition of multilocus linkage disequilibrium. In elementary algebra we have the analogous problem of defining the algebraic expression x^n when $n = 0$ (Lakoff and Núñez 2000). Note that our definition of a locus encompasses protein-coding loci, quantitative trait loci, and even single nucleotides.

Following Hastings (1984), the formulas for two- and three-locus multilocus linkage disequilibrium, in which $D_1(A_{k(i)})$ was substituted for $\Pr(A_{k(i)})$, are defined as

$$D_2 = \Pr(A_{k(1)}A_{k(2)}) - D_1(A_{k(1)}) \cdot D_1(A_{k(2)})$$

$$D_3 = \Pr(A_{k(1)}A_{k(2)}A_{k(3)}) - D_1(A_{k(1)}) \cdot D_1(A_{k(2)}) \cdot D_1(A_{k(3)}) - D_1(A_{k(1)}) \cdot D_2(A_{k(2)}A_{k(3)}) - D_1(A_{k(2)}) \cdot D_2(A_{k(1)}A_{k(3)}) - D_1(A_{k(3)}) \cdot D_2(A_{k(1)}A_{k(2)}).$$

Let D_n be the coefficient of linkage disequilibrium between n loci. Then the pattern here is that $D_n = \Pr(A_{k(1)}A_{k(2)} \ldots A_{k(n)})$ minus all possible products of lower-order linkage disequilibrium coefficients, such that each term has all of its subscripts adding up to n. The key to writing down an explicit formula for D_n is that the phrase "all possibilities of the subscripts adding up to n" refers to partitions of the positive integer n (Andrews 1976). A partition π of a positive integer n is a set of positive integers that adds up to n; *i.e.*, $\pi = \{n_1, n_2, \ldots, n_m\}$ such that $\sum_{i=1}^{m} n_i = n$. The set of all partitions of n is designated $p(n)$; *e.g.*, $p(5) = \{\{5\}, \{4, 1\}, \{3, 2\}, \{2, 2, 1\}, \{3, 1, 1\}, \{2, 1, 1, 1\}, \{1, 1, 1, 1, 1\}\}$. To define multilocus linkage disequilibrium, we have to add over all partitions, excluding the trivial partition $\pi = \{n\}$, and permute over all alleles for a given number of loci. However, the order of elements of the partition matters, and hence we construct the number-theoretic compositions c of the positive integer n (Andrews 1976). For example, all of the compositions for the partition $\pi = \{2, 2, 1\}$ are the ordered triples (2, 2, 1), (2, 1, 2), and (1, 2, 2). Using these mathematical notions we can generalize the two- and three-locus cases to define linkage disequilibrium between n loci as

$$D_n(A_{k(1)}, A_{k(2)}, \ldots A_{k(n)}) = \Pr(A_{k(1)}A_{k(2)} \ldots A_{k(n)}) - \sum_{\substack{\text{all compositions } c \text{ of } n \\ \text{except } c=(n)}} \left[\prod_{n_i \in c} D_{n_i}(\ldots)\right], \tag{1a}$$

where $n_i \in c$ means that $n_i \in c$ is a scalar component of the vector c. Equivalently,

$$D_n(A_{k(1)}A_{k(2)} \ldots A_{k(n)}) = \Pr(A_{k(1)}A_{k(2)} \ldots A_{k(n)}) - \sum_{\substack{\sum_{i=1}^{m} n_i = n \\ 1 \le n_i < n \\ 1 \le m \le n}} \left[\prod_{i=1}^{m} D_{n_i}(\ldots)\right]. \tag{1b}$$

The only way to decompose n into a single positive integer is $c = (n)$. Therefore, we can also write the highest-order coefficient of linkage disequilibrium as $D_n(A_{k(1)}, A_{k(2)}, \ldots A_{k(n)}) = \Sigma_{c=(n)} [\Pi_{n_i \in c} D_{n_i}(\ldots)]$, where the summation has only a single term and the product has only a single factor. Therefore, Equation 1a yields

$$\Pr(A_{k(1)}, A_{k(2)}, \ldots A_{k(n)}) = \sum_{\text{all compositions } c \text{ of } n} \left[\prod_{n_i \in c} D_{n_i}(\ldots)\right], \tag{2}$$

which we use below.

Equation 1 has never been written explicitly for general multilocus linkage disequilibrium, even though special cases have been given by Geiringer (1944), Bennett (1954), and Hastings (1984). The only explicit definition previously given for multilocus linkage disequilibrium is due to Dausset *et al.* (1978),

$$\mathbf{D}_n(A_{k(1)}A_{k(2)} \ldots A_{k(n)}) = \Pr(A_{k(1)}A_{k(2)}A_{k(3)} \ldots A_{k(n)}) - \prod_{i=1}^{n} D_1(A_{k(i)}), \tag{3}$$

which we call total linkage disequilibrium, $\mathbf{D}_n$, where we have again replaced $\Pr(A_{k(i)})$ with $D_1(A_{k(i)})$. We refer to $\mathbf{D}_n$ as total linkage disequilibrium because, as we show below, all of the nonboldface linkage disequilibrium coefficients $D_1, D_2, D_3, \ldots, D_n$ can be independent from one another and contribute to $\mathbf{D}_n$. Equation 3 has a simple heuristic interpretation: $\mathbf{D}_n(A_{k(1)} \ldots A_{k(n)})$ measures how far the haploid genotype at all n loci deviates from probabilistic independence.

We are now ready to derive the relationship between D_n and $\mathbf{D}_n$. In Equation 3, substitute $\Sigma_{\text{all compositions } c \text{ of } n} [\Pi_{n_i \in c} D_{n_i}(\ldots)]$ for $\Pr(A_{k(1)}, \ldots A_{k(n)})$ (see Equation 2), yielding $\mathbf{D}_n(A_{k(1)}, A_{k(2)}, \ldots A_{k(n)}) = \Sigma_{\text{all compositions } c \text{ of } n} [\Pi_{n_i \in c} D_{n_i}(\ldots)] - \Pi_{i=1}^{n} D_1(A_{k(i)})$. The last term in this equation is simply the value of $\Pi_{n_i \in c} D_{n_i}(\ldots)$ for the composition $c = (1, 1, 1, \ldots, 1)$, *i.e.*,

$$n = \underbrace{1 + 1 + \ldots + 1}_{n \text{ times}}.$$

Therefore, Equation 3 becomes

$$\mathbf{D}_n(A_{k(1)}, A_{k(2)}, \ldots A_{k(n)}) = \sum_{\substack{\text{all compositions } c \text{ of } n \\ \text{except } c=(1,1,1,\ldots,1)}} \left[\prod_{n_i \in c} D_{n_i}(\ldots)\right] \quad (4a)$$

or, equivalently,

$$\mathbf{D}_n(A_{k(1)}, A_{k(2)}, \ldots A_{k(n)}) = \sum_{\substack{\sum_{i=1}^{m} n_i = n \\ 1 \le n_i < n \\ 1 \le m < n}} \left[\prod_{i=1}^{m} D_{n_i}(\ldots)\right]. \quad (4b)$$

Equation 4 provides the crucial link between deviations from independence ($\mathbf{D}_n$) and the linkage disequilibrium coefficients D_n computed by GEIRINGER (1944) and her intellectual successors by decomposing $\mathbf{D}_n$ into the terms D_{n_i}, where $\Sigma n_i = n$.

DISCUSSION

We have converted the genetics problem of decomposing linkage disequilibrium into the mathematical problem of decomposing positive integers into their additive parts, all while maintaining the convenient heuristic definition of total linkage disequilibrium as the deviation from independence. Unlike GEIRINGER (1944), we can write down an explicit formula for multilocus linkage disequilibrium because we invoke partitions of integers and define $D_1(A) = \Pr(A)$, thereby merging her notion of linkage disequilibrium with those of DAUSSET *et al.* (1978).

One immediate consequence of our decomposition approach is that the single highest-order coefficient of linkage disequilibrium, D_n, cannot be examined in isolation. Because $D_n(A_{k(1)}, A_{k(2)}, \ldots A_{k(n)}) = \Sigma_{\text{except } c=(1,1,1,\ldots,1)}^{\text{all compositions } c \text{ of } n} [\Pi_{n_i \in c} D_{n_i}(\ldots)]$, we need to examine all lower-order linkage disequilibrium coefficients, $D_{n_i}(\ldots)$ with $n_i < n$. All of the subscripted linkage disequilibrium coefficients $D_1, D_2, D_3, \ldots, D_n$ can be independent from one another and all contribute to $\mathbf{D}_n$, which we therefore call total linkage disequilibrium.

Multilocus definitions of linkage disequilibrium have not been used very often in empirical studies because of the large number of inputs and linkage disequilibrium coefficients that must be analyzed ($2^n - 1$). Currently, even third-order linkage disequilibrium is seldom measured (THOMSON and BAUR 1984). However, explicit terms for multilocus linkage disequilibrium are of theoretical importance.

One important theoretical application is the analysis of multilocus epistasis. CHEVERUD and ROUTMAN (1995) developed a two-locus model of physiological epistasis that has been further refined by WAGNER *et al.* (1998). To analyze the evolutionary consequences of epistasis in these models, one has to first define linkage disequilibrium for a subset of the loci. Thus, to extend models of physiological epistasis to multiple loci, we must first define linkage disequilibrium for that subset of loci, which we have just done. Models of multilocus epistasis will be crucial in debates over what factors maintain coadapted gene complexes, increase additive genetic variance, and foster speciation (GOODNIGHT 1988, 1995; WADE and GOODNIGHT 1998).

We thank Phil Hedrick, Tom Dowling, and two anonymous reviewers for their helpful comments.

LITERATURE CITED

ANDREWS, G. E., 1976 *The Theory of Partitions.* Addison-Wesley, Reading, MA.

ARDLIE, K. G., L. KRUGLYAK and M. SEIELSTAD, 2002 Patterns of linkage disequilibrium in the human genome. Nat. Rev. Genet. **3:** 299–309.

BENNETT, J. H., 1954 On the theory of random mating. Ann. Eugen. **18:** 311–317.

BULMER, M. G., 1980 *The Mathematical Theory of Quantitative Genetics.* Clarendon Press, Oxford.

CHEVERUD, J. M., and E. J. ROUTMAN, 1995 Epistasis and its contribution to genetic variance components. Genetics **139:** 1455–1461.

DAUSSET, J., L. LEGRAND, V. LEPAGE, L. CONTÚ, A. MARCELLI-BARGE *et al.*, 1978 A haplotype study of HLA complex with special reference to HLA-DR series and to Bf, C2 and glyoxalase I polymorphisms. Tissue Antigens **12:** 297–307.

DOBZHANSKY, T., 1948 Genetics of natural populations. XVIII. Experiments on chromosomes of *Drosophila pseudoobscura* from different geographical regions. Genetics **33:** 797–808.

GEIRINGER, H., 1944 On the probability theory of linkage in Mendelian heredity. Ann. Math. Stat. **15:** 25–57.

GOODNIGHT, C. J., 1988 Epistasis and the effect of founder events on the additive genetic variance. Evolution **42:** 441–454.

GOODNIGHT, C. J., 1995 Epistasis and the increase in additive genetic variance: implications for phase 1 of Wright's shifting balance process. Evolution **49:** 502–511.

HASTINGS, A., 1984 Linkage disequilibrium, selection and recombination at three loci. Genetics **106:** 153–164.

HEDRICK, P. W., 2000 *Genetics of Populations,* Ed. 2. Jones & Bartlett, Sudbury, MA.

LAKOFF, G., and R. E. NÚÑEZ, 2000 *Where Mathematics Comes From.* Basic Books, New York.

LEWONTIN, R. C., 1974 *The Genetic Basis of Evolutionary Change.* Columbia University Press, New York.

LYNCH, M., 1991 The genetic interpretation of inbreeding depression and outbreeding depression. Evolution **45:** 622–629.

PALOPOLI, M. F., and C.-I WU, 1996 Rapid evolution of a coadapted gene complex: evidence from the segregation distorter (SD) system of meiotic drive in *Drosophila melanogaster.* Genetics **143:** 1675–1688.

RAWSON, P. D., and R. S. BURTON, 2002 Functional coadaptation between cytochrome c and cytochrome c oxidase within allopatric populations of a marine copepod. Proc. Natl. Acad. Sci. USA **99:** 12955–12958.

THOMSON, G., and M. P. BAUR, 1984 Third order linkage disequilibrium. Tissue Antigens **24:** 250–255.

WADE, M. J., and C. J. GOODNIGHT, 1998 Perspective: the theories of Fisher and Wright in the context of metapopulations: when nature does many small experiments. Evolution **52:** 1537–1553.

WAGNER, G. P., and M. D. LAUBICHLER, 2000 Character identification in evolutionary biology: the role of the organism. Theory Biosci. **119:** 20–40.

WAGNER, G. P., M. D. LAUBICHLER and H. BAGHERI-CHAICHIAN, 1998 Genetic measurement theory of epistatic effects. Genetica **102/103:** 569–580.

WRIGHT, S., 1931 Evolution in Mendelian populations. Genetics **16:** 97–159.

Communicating editor: M. W. FELDMAN