Table of Contents

Abstract

The evolution of sex is almost invariably viewed from the perspective of either the diploid 'individual' or a haploid locus. From those stances, the familiar costs, up to ‘twofold’, inevitably arise. These are traditionally assumed to demand a cryptic benefit, which remains elusive. This paper argues that these costs are not encountered when the perspective of the haploid genome is taken. In doing this, one can assemble a plausible evolutionary trajectory for a sexual system from its inception to the present without encountering any of the assumed costs.

Summary

Sex is still widely regarded as an unsolved evolutionary mystery. This is due in significant part to its behaviour in model systems, when pitted against perpetually asexual diploid lineages, and yet it is the dominant mode among many eukaryotic groups.

In eukaryotes, the central process of sex is periodic syngamy and reduction, giving alternation of haploidy and diploidy, generally with biparental inheritance. Given the apparent numerical superiority of asexual modes that avoid both pairing and reduction, most treatments either assume or conclude that sex must offer an adaptive benefit, when compared to asexuality, in order to explain its frequency among extant eukaryotes. Yet the nature of this benefit has so far proved elusive, even some 45 years after the landmark writings of Williams, Maynard Smith, Bell and Hamilton first articulated the difficulties of fitting sex into the framework of late-20th Century evolutionary biology.

In this essay, I argue that a definitive adaptive benefit of sex is elusive because the adaptive forces it is invoked to oppose are themselves illusory. The mystery of sex derives fundamentally from the wrong choice of perspective for analysis: typically that of the diploid organism or the individual locus. Taking either of these perspectives leads directly to a twofold error. If, as is probable, the transaction was initiated by haploid genomes, at no point in subsequent lineage history does the diploid genome take over as a fundamental unit while intermittent meiosis remains. Conversely, an individual locus can have no ‘interests’ without it, since that metaphor depends upon sex.

The currency of a recombining population is the haploid genome. From the haploid perspective, the transaction is one of temporary union, not puzzling division. From such a starting point, the complex asymmetries of modern sex can all be reached without ever approaching a point at which any cost becomes twofold, either suddenly or incrementally. The genetic transaction remains symmetric and mutually beneficial to both haploid partners, despite many layers of modern complexity. These complexities are themselves partially driven by the dynamic between these partners of approximately equal strength. When locked in permanent harness in an asexual diploid, meanwhile, a different and more damaging dynamic emerges.

The entire eukaryote clade coalesces upon sexual organisms; all asexual eukaryotes are secondary reversions. The spottiness and apparent youth of such asexual lines

suggest a disproportionate rate of extinction and cladogenesis between sexual and asexual eukaryotes, not a cryptic adaptive benefit to sex within each separate population. A careful examination of a plausible evolutionary history shows that a eukaryotic clade predominately consisting of sexual members can arise without a substantial threat from derived asexual diploid lineages. Taking this perspective, there is no requirement for a cryptic benefit to offset the assumed costs.

The Evolution of Sex

Sex Defined

For the purposes of this note, I define sex as cyclic syngamy and reduction: the alternation of haploid and diploid phases. A number of mitotic generations, in either or both of the haploid and diploid phases, may intervene between these phase switches. Furthermore the reduction process may or may not result in genetic recombination, either passive (deriving simply from independent segregation of multiple chromosomes where n>1) or active, involving a process of reciprocal crossover between homologous chromosome pairs.

I explicitly reject the common practice of synonymising sex and recombination, whether prokaryotic or eukaryotic. Recombination between parental genomes is an almost universal feature of eukaryotic sex, and many of the enzymes involved in its operation have homologues in prokaryotes performing similar basic functions. Recombination has such important consequences that it seems almost heresy to say that it is not the defining feature of sex. Nonetheless, the central transaction of eukaryotic sex is cyclic alternation of ploidy, not reciprocal exchange between homologous chromosomes during meiosis. The latter cannot even take place without syngamy and reduction.

Alleles and Genes

Sex has important consequences for levels of selection, and so it is important to be clear about the units of selection one considers on either side of the sex/no-sex boundary. Some confusion arises from export of subgenome levels of selection across the very boundary that is responsible for them.

In a mitotic lineage, whether haploid or diploid, entire encapsulated genomes form indivisible linkage units. Both segregation and recombination are absent, and so there are no genomic subdivisions qualifying for George Williams's definition of a gene: "That which segregates and recombines with appreciable frequency" (Williams, 1966). Being delimited by independent inheritance, such a unit need not map closely upon the molecular biologist's gene - a genetic segment producing a particular discrete gene product - nor a geneticist's – the unit of heredity responsible for a distinct character state. To reduce confusion, I use the term 'evolutionary allele' to distinguish Williams's genes from the others. An *allele* is a variant of a particular genetic stretch of interest, as long or as short as necessary for discussion. The length of an e*volutionary* allele depends upon the extent to which it possesses evolutionary independence from other such linkage units. A *gene*, meanwhile, corresponds more closely to a region producing a particular biochemical product, delimited as such regardless of the genetic system of its possessor.

There is ample opportunity for terminological confusion at this boundary between recombinant and nonrecombinant genotypes. If one adopts Dawkins's stance (Dawkins, The Selfish Gene, 1976), which owes much to the work of Williams and Maynard Smith, any subdivision of a genome possessing evolutionary independence

from other such subdivisions can be viewed as a 'selfish gene', even though it may span many discrete coding sequences. Dawkins's gene is my evolutionary allele. Such a unit is selected for amendments that optimise along its entire length, dependent upon the extent to which it is separated from other such units over the generations. This is critically dependent upon sex and, within it, upon recombination and segregation of subgenome fragments. It is particularly important not to export the metaphorical wishes of such genes/evolutionary alleles across the boundary that is responsible for their very existence. Whatever levers may be imagined available to selfish genes to enhance their own transmission, abandoning the arena that gives them their independence is not one. The metaphor breaks down at this boundary.

Introduction

Meiosis is overwhelmingly the rule among extant Eukarya, particularly among multicellular species. Even those species not known to perform sex are diploid, and many possess functional homologues of meiosis- or syngamy-specific genes (Speijera, Lukešb, & Eliáš, 2015), which suggests recent sexual ancestry, if not current cryptic sex. Asexual lineages appear to lack evolutionary staying power, given the ease with which genes involved in meiosis can frequently be identified despite apparent current disuse. If they are truly asexual, they appear not to have been so for long.

This distribution indicates that sex is foundational to the entire eukaryote clade, and it is likely that all modern lineages trace back through cyclically haploid-diploid ancestors through the Last Eukaryote Common Ancestor (LECA) ultimately to a singular event of initiation of the transaction. Modern asexual haploid eukaryotes are unknown; all modern eukaryotes possess a diploid phase either occasionally, or perpetually.

Given that all of the diploidy found in Eukarya probably originated from syngamous haploidy, it is curious that treatments so frequently look at the matter from the perspective of the diploid. It is as if having once established diploidy by syngamy, it should become the 'new normal', and haploids should essentially cease to have separate existence and interests. Sex should only exist for so long as it takes to generate diploids, which should then give it up as embarrassing experimentation! Yet there is no fundamental reason for this expectation beyond historical and personal perspective. We ourselves are mainly diploid, as is most of the visible biosphere for by far the greater part of the life cycle. The haploid phase of our life cycles is seen as almost incidental – a means 'for' dispersal, or for generating novelty or variation. The tools of population genetics themselves were formulated mainly to deal with alleles passed around populations of diploid organisms, with gametes as mere vectors. Sex sits ill within this framework.

Much of the 'mystery of sex' arises from consideration of simple contests, either experimental or theoretical, between those diploid lineages that reduce to haploids, and other functionally identical lineages that do not. In simple competitions in isolation, sexual reproduction is inefficient in anisogamous species. A diploid female can generate offspring with none of the costs of a partner, or those presumed to arise from segregational and recombinational severance of positive epistasis. Such

offspring will have all of her genes, instead of 50%. This leads to an exponential doubling of descendants by contrast with a sexual line competing for the same resources. For the individual locus, meanwhile, residence in every offspring seems a better fate than a net chance to reside in only 50% of them. Despite this apparent numerical penalty, sex clearly does more than adequately in the wild. The focus of much analysis, particularly since 1970, has been to explain this apparent paradox.

August Weissmann was one of the first authors to attempt to find reasons for the prevalence of sexual reproduction. Weissmann placed primary emphasis upon the generation of variant diploid organisms:

"I hold that the deeper significance of every form of amphimixis, - whether occurring in conjugation, fertilization, or in any other way,—consists in the creation of that hereditary individual variability which is requisite for the operation of the process of selection, and which arises from the periodical mingling of two individually different hereditary substances." (Weissmann, 1891)

The role of meiotic recombination in these 'individually different' hereditary materials was at this time presumably unknown, although the link of sex to variation has been the focus of most subsequent theories as well.

With insights provided from genetic recombination, Fisher (Fisher R. , 1930) and Muller (Muller, 1932) contended that sex allowed a population to fix multiple advantageous alleles somewhat independently, increasing the rate of generation of beneficial phenotypes. Muller further was responsible for the observation that recombination increased the capacity for a population to purge deleterious mutations that might otherwise accumulate due to stochastic losses of the least mutated genomes. (Muller, 1964)

However, all of these mechanisms require a variant population to build in the first place for their effect to become apparent. They cannot be, whether singly or collectively, a fundamental driver for the initial evolution of a complex system of syngamy and subsequent meiosis with the added layer of an equally complex system of reciprocal exchange, simply for 'the good of the species'. While they may help explain the evolutionary resilience of sexual lines, they cannot explain their origin.

This is the problem of group selection (Wynne-Edwards, 1962). With the recognition of multiple levels of selection, group selection received some criticism (Williams, 1966). Firstly, even if a characteristic benefits a population, it must still get from singular origin to widespread incidence by some means other than its ultimate benefit. More seriously, that group benefit could not evolve, or be sustained, if it proved individually costly. Sex is a common feature whose apparent benefits were at the group (and future) levels, but which appeared costly to individuals - 'the individual' being the diploid organism. If organisms, particularly females, are selected to maximise their reproductive output, then clearly halving it goes against this expectation.

Meanwhile, from the perspective of an individual gene (another perspective that became popular at about the same time), the problem remains – a gene that only gets

into 50% of offspring would seem naively to be selected to get into all of them if it could.

Williams (Williams, 1975), Maynard Smith (Smith, 1978) and Hamilton (Hamilton, 1998), progenitors of this 'Selfish Gene' perspective popularised by Dawkins (Dawkins, The Selfish Gene, 1976), all puzzled at length as to how a gene which promoted asexuality, and hence managed to get into every offspring, did not in the general case lead to the flame of sexuality being extinguished at the very outset. A gene that only gets into 50% of gametes was supposed by Williams to suffer a 'cost of meiosis'. Maynard Smith meanwhile noted that a female who produced only female offspring would have twice as many grandchildren as one producing males and females in equal number – the 'cost of males'. Further apparent costs relating to recombination, segregation and limitation by mates were also identified, but 'twofold cost' and 'sex' have become particularly closely linked.

The Costs

The famous 'twofold cost' is expressed either as a cost of producing non-productive offspring (cost of males or cost of anisogamy) or of dividing individual reproductive output into gametes (cost of meiosis). The cost of meiosis may be expressed from the diploid's perspective – the cost of 'halving' reproductive output - or from a 'gene's eye view' at a heterozygous locus – the cost of only getting into half of gametes. The costs of males and meiosis are often presented as two sides of the same coin, but as Maynard Smith himself pointed out (Smith, 1978), they cannot be equivalent since the twofold cost is absent in isogamous organisms, which have no males yet do perform meiosis.

The Cost of Males

In a species that concentrates offspring investment asymmetrically between two genders, we need not doubt that an asexual 'female' population (the producer of the larger of the two gamete types) could quite easily replace a typical dioecious one given a relatively homogeneous environment and genetic background, and adequate dispersion. For a female, male offspring limit her production of descendants, which would double its exponent if she only bore daughters. What is less clear is that this amounts to a cost of *sex*. A cost of *dioecious* sex, perhaps, but even this seems debateable. The essence of the gender distinction is differential gamete size, not dioecy *per se*. There is no parallel difference in genetic futures for the majority of genes contained within these asymmetric packages. Although possessing different amounts of cytoplasm, they possess approximately the same amount of DNA. Most genes occupy autosomal loci. Where unequal phenotypic cost occurs in a sexually dimorphic species, autosomal chromosomes possess genes investing in both strategies. The net return for each haploid gene set, and each gene within such a set, is equalised by approximately equal residence time in each type, averaged over multiple generations.

Male gametes are therefore not costly *to genes* if the net return from alternation between the two states is at least equal to that from perpetual isogamy. From the perspective of the individual autosomal haploid gene, sex has the same net effect

regardless of whether it passes through isogamous or through randomly alternating anisogamous packages. Nothing changes, from such a perspective, as we move from an ancestral state of residence in equal-sized packages through a gradual increase in package asymmetry, nor when multicellular bodies appear specialising in production of one gamete type or the other. The twofold cost of anisogamy does not appear at a sharp boundary; no fractional change in asymmetry leads to a sudden doubling of cost. But nor does it appear gradually, increasing by small increments with each incremental change in the relative investment in cytoplasm.

This is not to say that an asexual clone *cannot* extinguish an ancestral dioecious sexual species due to increased production of the productive gender. But the expectation that it *must* derives fundamentally from invalid assumptions regarding their respective and collective dynamics. When the sexual and asexual forms are lumped into one population and treated as a single unit with comparable dynamics, we miss the fact that *sex itself* both draws the boundaries of the sexual population, and provides the stirring that underlies the assumption of panmixis – random mating - implicit in the majority of algebraic representations. An asexual subpopulation lacks the vector of stirring provided by reduction to motile gametes and by mate search. It also invariably finds itself in competition with a resident sexual species, varied and locally adapted. The newcomer is free to extend beyond the range to which the sexual version is restricted, and increased production and density independence may help it to do so. But as regards *elimination* of the resident parent species, it finds a variant and locally adapted competitor, which it must defeat armed only with a clonal version of one of the ancestral population's genomes.

Such secondary diploid asexual reversion could be usefully analogised as a kind of 'species cancer': diploid overproduction, rather than an adaptation as such. As with cancer in an individual, the parent species may or may not recover from any given outbreak of asexuality. In this light, the use of efficiently 'stirred' population genetic models with minimal overall variation, developed mainly for use within sexual populations, is comparable to always assuming that metastasis will occur in cancer. This gives a false picture of the prognosis. The existence of cancer does not lead us to puzzle over the continued existence of living individuals.

The Cost of Anisogamy

Frequently, the cost of males is given an alternative name, the cost of anisogamy. It sounds more formal. These are not quite the same thing, however. The cost of males relates specifically to a dioecious species – one in which male and female gamete-producing organs are borne on separate individuals. Anisogamy relates to the difference in size of male and female gametes. In a hermaphrodite, the same individual can produce male and female gametes. The manner discussed above in which a constitutionally 'female' lineage can spread through the elimination of male offspring is not available to a hermaphrodite. The cost of anisogamy is therefore not a synonym for the cost of males.

The Cost of Meiosis

At the genetic level, again we might (influenced by Selfish Gene thinking[1]) agree that an allele might 'prefer' to get into every gamete instead of every other one. But if homozygous, as are the majority of loci presented to meiosis, it does so anyway. Heterozygous alleles, meanwhile, are generally competing for the gamete against variants of themselves. Insofar as they can succeed in distorting transmission in their favour, they may become common, but then homozygosity will increase. The metaphorical 'desire' of a gene to spread is an expression of the conventional mechanism by which it does so within recombining populations: by promoting, or minimally undermining, fitness via phenotype, and subsequently by integrating into the recombinant genomes of descendants. The opportunity presented by physical proximity in the diploid certainly does select for perturbations from equilibrium – for example 'drive', which alters the Mendelian 50% chance offered to the majority of genes. But such mechanisms *depend on* meiosis, they do not select for its abandonment.

The opportunity for transmission distortion is mechanistically available to only a small fraction of the genome, remains under the scrutiny of selection on organismal fitness, and is typically dissipative due to its interaction with heterozygosity. It is a common error to equate the summed metaphorical wish of haploid genes to defeat their alleles with pressure toward the complete abandonment of gametes. Only through sex can individual subgenome fragments progress through the population. Only through sex do fragments of the genome have 'desires'. It is something of a paradox from the gene's eye view - whatever interests it might be seen to pursue as an evolutionary unit, one that it cannot pursue is one in which it stops being such a unit, subsumed into a frozen genome.

While outcrossing sex remains in place, there is some squabbling in the 'parliament of genes' between members on opposing benches, but the votes at each locus cancel out.

The Costs of Recombination/Segregation.

New genetic combinations are generated by eukaryotic sex from diploid genomes in three ways, all depending upon the fact that segments of that genome can follow independent trajectories through lineages:

a) Independent segregation/syngamy of homologous chromosomes
b) Independent segregation/syngamy of nonhomologous chromosomes
c) Intrachromosomal crossover

Such recombination has two fundamental potentially detrimental consequences, and hence may be a separate source of cost for sex:

1. Loss of adaptive gene combinations promoted by selection
2. Mutagenesis (restricted to c above)

Nonetheless, there are offsetting benefits. The immediate benefit is cytological. At least one chiasma per chromosome pair appears to be necessary for proper disjunction at the metaphase plate, through effects on equational tensioning of opposing spindles and on polar orientation prior to cytokinesis. One might wonder why meiosis occurs

[1] Dawkins (Dawkins, 1982) portrays chromosomes being 'dragged kicking and screaming into the second anaphase of meiosis'!

at all, but given that it does, avoiding aneuploidy through crossover is beneficial. This would render population genetic considerations on the 'Cost of Recombination' somewhat moot – its retention merely requires that the cost of aneuploidy be greater than the longer-term population cost of crossover, when comparing a subpopulation that performs it to an otherwise equivalent one that does not. This would be a difficult comparison to perform empirically, since crossover is deeply embedded in modern meiosis and cannot be disabled without other consequences. Nonetheless, those who insist that recombination *is* costly, simply by analysis of its promotion of linkage equilibrium among adaptive combinations, are themselves doing so without accurate assessment of the *overall* net cost, taking all consequences into account. The 'mystery of sex' here derives simply from the assumption that this is negative in sum, not from any empirical fact.

Population effects due to crossover cannot be ignored. But analysis often takes a rather one-sided view. The perceived detriment is of breaking beneficial associations. But a significant source of such beneficial combinations is recombination itself, affecting not merely two linked model loci, but all interactions between all loci occurring either side of a given break. Whether breakage of a particular beneficial linkage A-B takes place depends not just upon the placement of the crossover between them, but on the composition of the homologue. While A-B is rare, a positive epistasis may indeed be broken, but this does not take place in every chromosome in every generation. Selection merely needs to promote copies of the association at a greater rate than such losses occur for A-B to increase. In fact, the frequency of alleles A-x, x-B *and* A-B will all increase in the population if this condition is met. This progressively reduces the chance of A-B severance on those occasions when crossover does occur between the loci on an A-B chromosome.

Therefore, the circumstance in which recombination is most heavily penalised is in its occasional breakage of rare beneficial linkages. This is hardly likely to be a major cost, and is offset by its effect on the initial generation of such linkage groups that are subsequently promoted by selection.

Conversely, any negative epistasis C-D can also be broken by crossover, which in tandem with selection will cause decrease in the frequency of C-D individuals, and the relative enrichment of genomes in alleles that do not participate in such negative interactions. There are, in reality, many pairs A-B and many pairs C-D appearing on either side of a particular crossover site. It cannot be robustly concluded that selection must always act to eliminate crossover entirely based only on negative consequences.

To summarise: taken as a whole, the population effect of crossover comprises:

- Non-interacting alleles:
    - Generates mosaics of beneficial alleles of independent origin
    - Purges individual detrimental alleles without interference
- Interacting alleles:
    - Creates positive (and negative) associations
    - Breaks negative (and positive) associations.

It is not clear that the occasional parenthesised consequences that go 'the wrong way' in the second list constitute a *net* overall cost of crossover, given that those which go 'the right way' would not occur in its absence.

The cost of mates

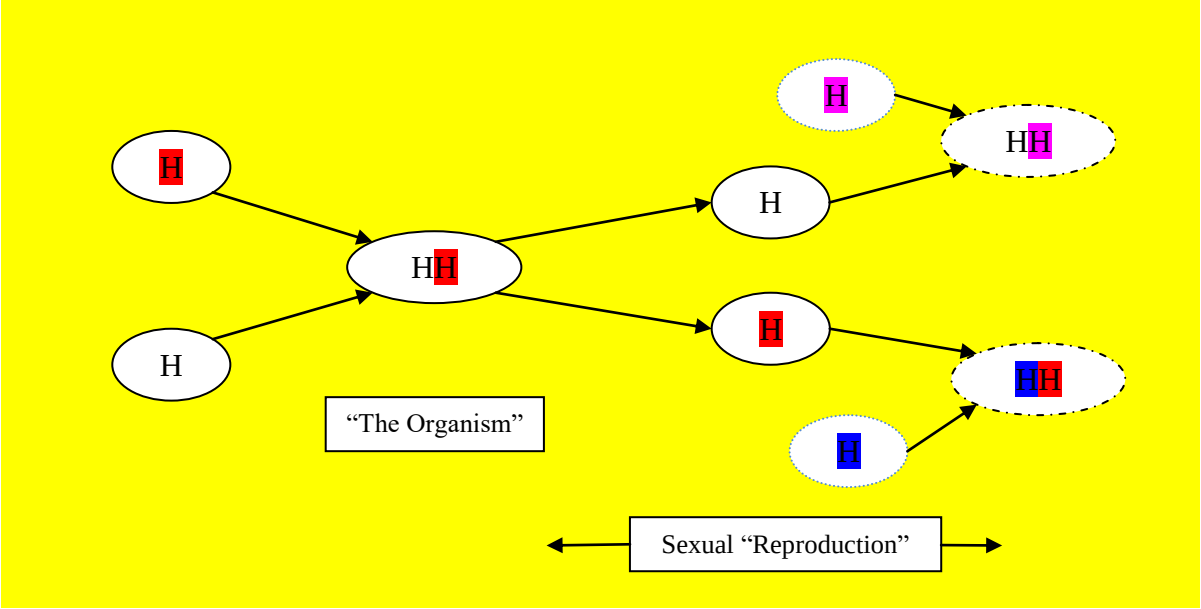
To undergo any kind of sexual cycle, it is necessary that gametes periodically locate each other. This cost is most significant in an obligate sexual species, which could conceivably be replaced by an equivalent asexual when sexual individuals are at low density - although the probability of such a lineage arising is correspondingly low, due to the relationship between mutation frequency and local population size.

Mate/gamete location is not such a penalty for a sexual species that retains the parallel capacity to reproduce by mitosis, however. Opportunistic matings could be performed by such a species with little obvious disadvantage compared to a purely clonal competitor.

The extent of this potential cost depends very much upon the circumstances of the species – whether multicellular or not, dioecious or not, outcrossing or not, mitotically competent or not, and on whether gamete dispersal and mating have any additional life-cycle consequences beyond the simple requirements of reproduction. Certainly, this factor cannot be regarded as a *general* cost of sex – a contributor to mystification at its very existence throughout the eukaryote tree.

The Sexual Transaction

There are numerous variations on the fundamental theme of cyclic ploidy alternation. They can all be stripped away: the consistent beating heart of 'minimal sex' is a cycle which is properly represented as *beginning* with a haploid possessing a single chromosome H pairing with another, merging their cytoplasm and possibly (but not necessarily) their nuclei, later followed by separation back to the haploid state:



The transaction depicted above is isogamous – there is no differential in relative gamete size. As is well known, there is no inherent cost in this – to the haploids emerging, there can be no intrinsic extra cost arising from fusion/reduction by comparison to simply doing nothing. The merged state is misleadingly intimate – for a greater or lesser time, two genomes occupy the same compartment. But, for all the difference it makes, one could simply eliminate all circles 'HH' and extend the arrows across the resulting void – ships that pass in the night. HH – the diploid – is a transient stage, not an entity from whose perspective the entire transaction would be viewed. That stance is something of a diploid prejudice.

This picture is deliberately naive. I have omitted

- Germline mitosis, which can occur in either or both of haploid and diploid phases, depending on species.
- Multicellular somatic cell lines derived mitotically from, according to species, haploid or diploid cells or both.
- Gender differences in gamete size and packaging.
- The serial production of diploid organisms in certain species without intervening meioses.
- The fact that reduction in modern organisms is almost invariably preceded by a doubling step, to give four haploid outputs.
- Recombination, either coarsely achieved through independently segregating chromosomes or finely through reciprocal crossover.

These muddy the waters when it comes to considering the central sexual transaction, that symmetrical contract between two haploid genomes.

This cartoon version of the cycle is readily reducible to a primordial state. The decision to start with the haploid egg, not the diploid chicken, is a deliberate one, and is quite fundamental. There are no significant costs associated with such a symmetrical transaction. But note that if, further, the *genetic* transaction remains symmetrical from the perspective of each gene, then there appears to be no means by which additional genetic cost can arise if the cycle adds phenotypic asymmetry and complexity. Approximate 50/50 impartiality of genetic outcomes is maintained throughout the sexual world. Like Einstein's falling roofer, the haploid 'feels no force' while net genetic symmetry prevails – from the inside, the transaction looks the same regardless of the asymmetry of packaging or investment.

To expand this theme, I will take a progressive approach, commencing with a plausible initial state and discussing the elaborations and the relevance of cost analysis, in approximate evolutionary sequence. During this progression, two related questions are considered:

1) What, if anything, can truly be said to suffer a cost from sex?
2) How might that cost manifest itself, as a brake upon the evolution of a given innovation?

Origins

As mentioned, it seems to be the case that sex is a synapomorphy in eukaryotes: the entire eukaryotic clade appears to share common ancestry with a single sexual species. Whatever else there was during the transition from prokaryotes to eukaryotes, it has left no descendants other than through the surviving eukaryote lineage. This gathers into a coalescent 'singularity' many innovations presumably acquired during prior history. Because of this, the order of acquisition of these features is uncertain. However, it is certainly a reasonable supposition that sex originated in an organism furnished with mitochondria, a linear genome, a cytoskeleton and mitosis. It is probable that many mitotically reproduced eukaryote lineages will have existed at this time, but they were likely all haploid. No exclusively haploid cell lines exist today among eukaryotes. Yet at some stage of evolving from prokaryotes to eukaryotes, diploidy must have arisen. The first question to address, then, is whether diploidy is likely to have preceded sex (as defined), or been caused by it through syngamy.

Transient diploidy occurs during the replication of any haploid cell, including prokaryotes – sister chromosomes inevitably occupy the same cell prior to completion of cytokinesis or its equivalent. But we are here interested in a diploid *organism*: one in which the diploid state has additional life history significance beyond the inevitable transient diploidy of replication.

The haploid pairing in any diploid cell must comprise relatives, at some remove – a gene from either haploid set will share common ancestry with its corresponding locus on the other. Did that coalescent, in the first diploid to undergo reduction, occur in a

diploid ancestor (arising through endomitosis) or from independent haploid lineages (arising by complete mitosis)?

The first is unlikely to be a precursor of the haploid-diploid cycle. Although endomitotic diploids may occur, the subsequent ejection of a hopeful haploid pair from a diploid line originated in this way would find them with no viable partners other than each other.

This is likely to provide a general principle: *at any stage, sex cannot be achieved from the start point of a mitotically competent diploid.* Once lost, it cannot be regained.

At the origin, further obstacles exist. If the lineage has never previously performed syngamy, how likely is it to evolve it to coincide with reduction? Further, the likeliest benefit of endomitotic diploidy, retention of a closely related backup genome copy for repair through the G1 phase of the cell cycle, is directly undermined by reduction. It is noteworthy that, during G1 in modern diploids, repair favours the potentially mutagenic 'end-joining' pathways over the homologous repair pathways, despite the availability of a homologue. Homologous repair favours 'true' sisters: those generated during the current cell cycle by mitotic duplication.

On the other hand, difficulties are far fewer if we *start* with syngamy. A haploid-initiated fusion/separation system would enable the sexual experiment to be started by a single haploid individual, without commitment. In such a scenario, there is a freely available source of haploid partners for syngamy - relatives. Partners need not be willing – or expressed more formally, it is not necessary that the capacity to initiate fusion exists anywhere else the population before it can occur at all. A population of 'initiators' can build up, increased in frequency by any benefit that accrues in the diploid state.

'Syngamy-first' thus makes minimal demands upon probabilistic resources, and has two significant implications:

1) Sex is not 'for' diploid organisms, nor individual genetic loci, but for haploid genomes. They are the beneficiaries of temporary diploid union, and haploidy is properly viewed as the native state, not as a mysterious intermediate in an adaptation of diploids. The diploid is a transient binary phase in the life cycle of a haploid organism. This has not ceased to be true today - despite the many variations on the basic syngamy/reduction cycle, we remain binary organisms formed from temporary haploid alliances.

2) Sex is not 'for' recombination, at least not initially. Although many treat sex and recombination as virtual synonyms, eukaryote recombination cannot even occur until syngamy and reduction are in place. The lack of individual benefit, problematic for recombinational theories, can at least be deferred.

For such a tendency for temporary pairing to propagate, there must obviously be some benefit in doing so, but the benefit accrues to one or both of the haploid partners, not to the diploid that they form. A 'couple' is not a unit with interests, but two individuals associating for mutual benefit. Thus, when the partnership is severed, there is no twofold cost of halving – the transaction is properly rendered as

1+1 -> 2 -> 1+1

not

1 -> ½ + ½.

Nothing is lost. In a series of such unions and partings, the fundamental units of selection are the two haploid genomes, experiencing selection both singly and in pairs.

Syngamy

The challenge remains to provide both a plausible rationale for syngamy, and a reason why that diploid state did not become permanent. According to the foregoing reasoning regarding long-term diploidy, the sexual cycle demands diploid/haploid alternation almost from the outset. This requires a certain amount of speculation, given that we are dealing with an unknown organism in an unknown environment about 1.5-2 billion years ago.

Plausible general benefits of primitive syngamous diploidy for haploids include:

- Complementation of detrimental alleles in the diploid state.
- A rapid increase in size may provide some protection against a threat suffered by smaller, lone haploids.

Somewhat less plausible ideas, in my opinion, include:

- The repair hypothesis. This is dubious since a spare chromosome copy for mutational repair would be much more readily achieved by endomitosis than syngamy. This would be a more accurate backup, and would reduce the risk of harbouring deleterious recessives, which in the presence of gene conversions induced by the repair process itself form something of a ‘time bomb’ for a long term non-reducing diploid. But, as noted, endomitosis is not a plausible starting point for reduction. Indeed its very existence confounds the repair hypothesis. One would not discard a closely related backup for one of uncertain composition.
- ‘Selfish’ hypotheses, whereby a genetic element increased its spread by promoting syngamy purely to effect a vampirish conversion of uninfected units into versions of itself. This is self-limiting, in that conversion of the population rapidly consumes the substrate for such an element’s continuing drive, and hence is unlikely to be of long-term effect.
- Recombinational hypotheses. The main class of ‘theories of sex’, these assume that sex’s most significant feature is the reason for the whole enterprise, even though recombination is also noted as a cost of sex. A fundamental obstacle would appear to be the complexity of syngamy-plus-duplication-plus-recombination-plus-reduction, all required before any recombinational benefit (probably itself weak at the generational level) could be achieved.

Initially, we can only guess at a rationale, applying as it does to a minor species in a tiny corner of the globe. Without knowing its ecology or other constraints, we cannot be definitive about what benefits intermittent diploidy may have given it. But it is certainly not necessary that those benefits follow the same rationale in every species continuing to practice that life history today. A Grand Unified Theory of Sex is not required for this one species, and nor should a failure to provide one trouble us fatally at this stage – what is it battling? There is no twofold cost of males, there is no twofold cost of meiosis, and costs of segregation and 'mate search' will be minor at best. If cyclic syngamy arises within a population of such asexual haploids, it need provide only a relatively small benefit to prosper – complementation and size change are certainly plausible candidates. Reproductive competence would continue to be provided by haploid mitosis, even among those competent for syngamy. If such a coupling strategy can prosper against a background of ancestral non-syngamous haploids, by however small a benefit, then the threat from reversion to that state is negligible at this stage.

Dominance

The theory of dominance evolution has a history of controversy longer than that of sex, and modern dominance dynamics may bear no relation to primitive ones. However, it is clear that a haploid pairing such as I envisage will immediately encounter an environment where both alleles at a locus have the potential to be expressed. A diploid formed from very distantly related haploids may be less likely to be viable, but among closer relatives there would still be expected to be a significant proportion of heterozygotes. To be immediately beneficial for complementation, the dominance relations of these heterozygotes must go the 'right way' – most alleles deleterious in the diploid state must be recessive – immediately.

Where a locus is homozygous, it is likely that dosage mechanisms operating in the ancestral haploid will ensure that overexpression is suppressed. For more than half of the cell cycle, after all, chromosomes exist in pairs even in haploid lineages, and so mechanisms to deal with overexpression from multiple gene copies should already be available. For heterozygous loci, mechanisms may be a combination of complementation (where one allele's product is simply absent) and dominance (where partial or complete suppression of an allele in the presence of the other occurs). It remains unclear as to whether dominance is an evolved response or an inevitable consequence of gene action. Fisher favoured the former (Fisher R. A., 1928), Wright the latter (Wright, 1929). There may be some ground for both to be partially correct in modern diploidy, but we clearly can't appeal to evolution at the outset.

But there does appear to be a natural and immediate capability of heterozygous alleles to interact on phenotype with a dominance relation. In hybrids, where excessive genetic distance leads to infertility through failed meiosis, the organism itself is frequently not just viable but 'vigorous', despite the probability that many of the heterozygous alleles have never encountered each other before. This may be partly due to dominance of non-deleterious alleles. Therefore, it seems reasonable to expect a similar kind of vigour in early syngametes, and indeed this may be a key benefit of such early unions.

Reduction

Reduction must follow syngamy with some rapidity for alternation of ploidy to remain plausible. Arguments made above against the pre-sexual population consisting of endomitotic diploids apply equally to those of syngamous origin. There is a limited 'window of opportunity' for reduction to arise in a long-established diploid line for the following reasons:

- After divergence from/extinction of any strictly haploid ancestral lineages, reductive haploids would struggle to find compatible partners (other than each other).

- Complementation in an extended diploid phase allows for an accumulation of deleterious recessives, threatening the viability of future haploids, which would possess these recessives in single copy exposed to selection.

- Gene conversion, a by-product of homologous repair mechanisms, would lead to an increase in deleterious homozygous recessives, threatening the diploid lineage itself.

- Diploidy offers no selective protection to alleles expressed mainly in supporting the haploid state, so the haploids' chances of subsequent emergence or independent survival following extended diploidy would be expected to diminish progressively with disuse.

I would therefore dismiss, on these principles, the possibilities that syngamy and reduction arose a significant period apart. On the other hand, there appears at first glance no compelling reason, given any advantage to diploidy, for a return to haploidy.

It is however noteworthy that the machinery for reduction is readily available at the outset, in the final phases of mitosis, and it may conceivably be set in train by the very act of syngamy. Since the cell cycle is controlled by growth, syngamy might be expected to push the cell quite rapidly towards the division phase of mitosis. The cell is signalled that it has reached a point close to prophase: it has increased in size and has at least one chromosome pair, and would soon be triggered to separate them and divide. Therefore, it may not be necessary to invoke an adaptive explanation for reduction at all. The first innovation might initially be to *defer* division, not to invoke it.

Synapsis

In such a scenario, there is a requirement to explain the origin of synapsis. Co-location of sister chromatids in modern mitosis is not necessary, as the replicated chromatids are physically linked by cohesin and by the centromere. These provide both physical linkage and the sites of attachment of spindle fibres for segregation. Since we do not know the order of acquisition of centromeres and pro-meiotic reduction, it is not certain that the modern method of chromosome attachment and

severance was in place at this origin. Nonetheless, centromeres take part in segregation now, in mitosis and in meiosis I and II. It seems reasonable to assume that the transient endodiploids produced during haploid mitosis in this early species had sister chromatids in physical conjunction, whereas the syngamous haploid chromosomes envisaged in this discussion would be presented to this process physically separate.

However, from the point of view of the haploid, its only experience of diploid chromosomes comes from the sisters it houses in all but the G1 phase. The homologues from syngamy might resemble these sister chromosomes sufficiently to pass as sisters. Therefore, even if they are not quite presented in the physically attached form mitosis is used to, it has mechanisms to deal with aberrant sisters. The cell has numerous opportunities for error during its cycle, and various mechanisms ensure that these errors are fixed with sufficient fidelity. Chromatid pairs that are experimentally micro-manipulated out of alignment cause suspension of mitosis while they are shuttled back to the metaphase plate to try again. The presence of genes controlling these and other chromosome manipulations demonstrates their importance. Even though imperfections in mitosis can lead to defective products, they are evidently better than giving up, since the genes that provide rescue are so ubiquitous. There are therefore plausible mechanisms already embedded in mitosis that could deal with syngamously-sourced homologues as if they were errant, mechanically separated sister chromatids.

Physical links can also be established between two completely separate chromosomes. It is noteworthy that efficient chromosome pairing on homologous sequence is a fundamental requirement of homologous repair, a process likely to be substantially more ancient than sex and diploidy. Homology search in meiotic repair is mediated by sister strand invasion and the favourable energetics of complementary base pairing, a physical mechanism foreshadowing both mechanical synapsis and meiotic recombination, and hence available for non-sister conjunction in these early sexual unions.

There is thus no requirement to evolve a complex meiosis at the start before reduction can be effected – all of the basic toolkit is available in the mitotic cell cycle.

Cyclic Haploidy/Diploidy.

We now have possible rationales for a syngamous diploid phase, and a means by which reduction of such diploids may be simply achieved, which can cause a cycle to rumble into action. However, if diploidy is a beneficial state, we might expect selection to favour its permanent establishment to form a new class of mitotically competent diploids. This is the somewhat artificial dichotomy between the evolution of sex and its ‘maintenance’, as if it were a brief aberration that should have been abandoned the moment diploid mitosis could be achieved.

The following may act against this:

- Replicative competence in the diploid phase cannot be taken for granted. Analogous to the separation of algal and fungal components of lichens at

reproduction, the unit may have been be forced to return to the haploid, single-chromosome state in order to reproduce during early evolution.

- Depending upon the actual benefit of the diploid state, permanence may act to degrade that benefit. If, for example, the benefit comes from outcrossing heterozygosity, and homologous repairs are periodically triggered in the diploid (see below), they would have the effect, over time, of increasing the homozygosity of the haploid partners through gene conversion, dissipating that benefit as well as potentially unmasking genes deleterious in the homozygous state.

- The haploid state may retain a significant role in the life cycle beyond its mere existence as a transient vector for genes.

- There may be fluctuating effects of the environment upon relative selection in the two modes, such that the diploid is not continuously favoured. These need not be cyclic, but simply periodic.

The above are all speculative, contingent and not amenable to direct investigation in the actual organisms. They are suggested as potential mechanisms, not mutually exclusive, that would ensure that haploid genomes remained *periodically* produced without reduction to them requiring a specific adaptive explanation. At this stage, haploidy is the normal state, diploidy a temporary convenience (and long-term inconvenience).

Repair and Diploidy

During the ancient process of homologous DNA repair (HR), breaks in one chromosome may be repaired by copying a patch from a related chromosome across the break. If the original chromosomes differ in this region, the patch will result in a gene conversion event, where the gene copy on one chromosome becomes that on both. During mitosis, a sister strand is available, which is preferred, being an intact and near-identical version of the broken chromosome. However, a proportion of repairs use the homologue even when a sister is available. In a primitive system, with no prior diploid history, there are unlikely to be mechanisms capable of distinguishing homologues from sisters, and so the modern bias in favour of sisters is likely to be an evolved mechanism.

Following cytokinesis, the sister is no longer available anyway, being in a separate cell, and so the only options are using the homologue or performing end-joining, at the risk of loss of sequence. In modern diploid cells, end-joining is significantly preferred during the G1 phase (when a sister is absent but there is a homologue), suggesting that the mutagenic effect of end-joining gene loss is actually preferable to that of gene conversion-induced homozygosity. Again, this is likely to be an evolved response not available in primitive cells. Nonetheless, the fact that homologous repair appears to be a last resort even when a homologue is available argues against homologous repair being the principal driver of syngamy and diploidy.

This trend towards increasing homozygosity, at its strongest in the primitive system where the sister/homologue distinction is less readily made, is not in the long-term interests of the diploid lineage. Although there are possibilities for fitness enhancements to arise, more frequently gene conversion and repair-induced deletions on a diploid genome would be expected to be degradative. This is likely to be a fundamental mechanism limiting the long-term evolutionary persistence of diploid lineages. (Tucker, Ackerman, Eads, Sen Xu, & Lynch, 2013) 'Ancient' asexuals, most famously the rotifers, may have some means of opposing this effect, displaying widely divergent genomes in which diploid homology of some chromosomes is undetectable. (Welch & Meselson, 2000) . But for the general case, gene conversion is likely to be a significant threat to the long-term diploid.

Transposons and Diploidy

Transposons comprise sequences that are capable of copying or excision from one part of a genome and insertion into another. Many break upon landing, but while competent, a 'copy-paste' transposon has a substantial capacity to do damage, by disruption of coding or regulatory sequence, or by simply increasing the amount of material and energy required to replicate the genome. In an asexual lineage, active and damaging transposons tend to be self-limiting, tending towards eradication of the lineage in which they occur. With the intimacy afforded by syngamy, however, a mechanism is provided to allow transposons to move between lineages and remain in the population as potential infections at lower levels. Sex is thus somewhat unhygienic in that respect. A transposon may rapidly colonise a population of syngamising haploids. Those that it damages are removed, but it can remain active within the rest of this proto-population, in an uneasy stand-off between activity and selected mechanisms opposing this activity.

If a diploid, with a load of transposons that cause minimal damage in a sexual population, mutates to permanence, it has now imprisoned these mutagenic units. Although their infection of genomes is *caused* by sex – a kind of primitive sexually transmitted disease – abandonment of sex by diploids of sexual origin risks incorporating damaging transposons into a unit that is significantly less well equipped to deal with them. (Dolgin & Charlesworth, 2006) This further diminishes the threat of extinction of the sexual by asexual reversion from the diploid phase of the cycle.

Epistasis.

The selective effect of many alleles is influenced, positively and negatively, by the genetic contexts in which they find themselves - there is *epistasis* between loci, such that alleles effectively have different selection coefficients depending upon the alleles occupying other loci. Within the bounds of an evolutionary allele, these inter-locus effects are somewhat notional – the entire linked and/or encapsulated unit is only ever tested as a whole, and stands or falls by its net result over multiple trials. One could simply view the undivided unit as selection does: as an extensive allele whose composite selective advantage is a resultant of the combined selective effects of all its individual loci *and* any synergistic or antagonistic fitness interactions between them.

Epistasis has been suggested as a driver for the evolution of sex-with-recombination (Kondrashov, 1988). This, however, appeals mainly to intra-chromosome epistasis exposed by crossover, a subdivision not available at this proposed primitive stage. The role of such epistasis is unlikely to be great when we consider that selection on non-crossover chromosomes is blind to interactions within them. The population will, through whole-unit selection, contain fewer linkage units with negative intra-unit interactions and more with positive. For Kondrashov's scenario, one must presuppose a set of chromosomes somewhat crippled by negative epistasis but still common and evolving, discovering a means to separate these interacting loci to immediate benefit. It is not clear how such units would become common, and having reduced the load of negative epistasis, it is not clear what would then favour retention of the mechanism.

Segregation.

Strictly, segregation refers to the separation by meiosis of the alleles at a locus, a locus in a recombining population being coextensive with a protein coding gene, or any chromosome segment responsible for an equivalently discrete character state. There is no active recombination at this stage, so such units are not exposed individually. But since a 'locus' can really be as wide or as narrow as suits the discussion, I will include in this section both conventionally-defined allelic relations, restricted to discrete segmental states, and the broader collective that comprises the 'evolutionary allele'. In a segregating population without crossover, the evolutionary allele is an entire chromosome.

In the scenario being followed, segregation is merely an inevitable consequence of return of the constituent haploids to their prior, independent existence. But there is a dynamic not merely analogous, but homologous to the modern separation of diploid genomes into haploids. Perhaps surprisingly, the minimal process, even when recombination appears to be absent, contains all the significant features of modern sex as inevitable by-products of the cellular and population dynamics that result from ploidy alternation. It is not necessary to invoke specific benefits beyond explanations for syngamy and reduction already discussed, but the consequences are significant.

With only syngamy and segregation in place, three kinds of interaction can be contained within a diploid pairing:

- Heterozygosity or homozygosity of the alleles at a particular genetic locus.
- Epistasis between non-overlapping genetic loci on homologous chromosomes
- Epistasis between loci on nonhomologous chromosomes.

Where n=1, only the first two can be in operation.
Where n>1, all three can occur.

The first interaction has been proposed to lead to a 'segregation load'. If hybrid vigour – heterosis - is the reason syngamy spreads, then this implies a fitter phenotype for heterozygotes than either homozygote (or the 'naked' allele, in the free-living haploid). By breaking such heterozygotes, segregation acts against this benefit and, at least notionally, becomes a cost of sex. As has already been noted, however, if such genomes are arrived at by syngamy in the first place, it is hardly a 'cost' to deprecate

an indispensible part of the process by taking it in isolation. If the heterozygotes were arrived at by some means other than syngamy, then the stance might be more justifiable, but then the fitter genotype broken by sex appears to have been parachuted in from nowhere.

Gene conversion

Already discussed under 'Repair and Diploidy', gene conversion is an inevitable consequence of homologous repair pathways. It will be experienced during the diploid state by both transient (sexual) and permanent (asexual) diploids. This process has significantly different consequences in a diploid genome that remains undivided, compared to one which undergoes segregation. Conceptually, gene conversion lies somewhere between crossover and mutation in its effects. Such an event may be seen as a pair of 'mini-crossovers' occurring very close together on the chromosome, albeit without the reciprocal component of true crossover. Alternatively viewed, it has the effect of providing repeat mutation at a locus, with a greater bias away from detrimental mutations than 'true' mutation due to the effect of selection on the gene conversion donors present in the wider population.

Gene conversion has two principal effects:

- It increases homozygosity
- It generates novel composite genotypes by copying genes into other backgrounds.

The increase of homozygosity in a permanent diploid is one means by which beneficial recessives mentioned above may come to be expressed - one of the drawbacks of long-term diploidy is the masking effect of dominant alleles, which can slow down evolution by comparison to a species that indulges cycles of haploidy. But the downside of this process is that detrimental recessives can also come to be expressed as homozygotes by the same mechanism. Such recessives will come both from the original haploid partners, and from subsequent mutation, which has twice the target of a lone haploid. When a recessive mutation hits a haploid locus in a cyclic diploid, it is unlikely to stay in harness long enough for the change to ricochet onto its partner. But these changes will inevitably accumulate in a permanent diploid, as homozygosity increases, exposing it to a version of 'Muller's Ratchet'.

Muller noted two consequences of recombination: the combination of novel beneficial genotypes (also noted by Fisher) and the purging of deleterious mutations by swapping them out of otherwise advantageous contexts. These are somewhat equivalent (Felsenstein, 1974) – both viewpoints depend upon the reciprocal swapping of *relative advantage*, with selection favouring the composites where this has increased. In a cyclic diploid, repair-induced gene conversion can act in a similar manner to full-blown crossover. Occasionally, gene conversion will copy a deleterious recessive from one partner to the other. But just as often, a beneficial allele will obliterate the recessive. Selection will ensure that the results of the latter are found more frequently in the population than the former.

This mechanism is capable of incrementally enhancing the fitness of an evolutionary allele, even where it only replaces a detrimental allele with one slightly less so. Both increases and decreases of fitness will occur with approximately equal frequency, but

with the added bias of selection, this mechanism will raise the mean fitness of sexual genomes more rapidly than the asexuals, which must rely on mutation alone, and to which gene conversion is a ticking time bomb.

And hence, the population that continues to segregate its genome into haploids will be able to 'tune' its genomes more effectively to circumstances than the non-segregating diploids, simply by avoiding extensive liaisons. While the diploid lumbers on suffering the drag of a larger genome and negative heterozygote effects, the segregating population generates mosaic diploids from among the better haplotypes available, better not just in isolation but also in their capacity to interact successfully in diploids. Alleles tending towards negative interactions are purged more effectively; likewise, positive ones are promoted, and additional beneficial linkages are formed from a process that, in the permanent diploid, is detrimental.

There remains a degree of interference between loci within these broader evolutionary alleles, as there is no crossover-mediated recombination. Any advantageous allele is still linked to many partners that may be less so. But the evolutionary units are half the size or less of those in an asexual diploid, depending on the haploid chromosome number. Therefore, they behave as evolutionary alleles *sensu* Williams. Even with segregation alone and one chromosome, by small degrees the sexual clade can start to rumble into life, an accelerating set of populations with the capacity for subgenome tuning, although yet (if n=1) without genetic integration.

Recombination I

With more than one chromosome (which may be the case even at inception), we see this effect enhanced further – independent segregation of multiple chromosomes gives recombination, passively and unavoidably.

Most theories of sex are really theories of recombination. Many argue the processes to be essentially synonymous, hence the frequent use of the term 'bacterial sex'. However, the reciprocal recombination of eukaryotic sex is only distantly related to the various bacterial mechanisms, through features common to pathways of DNA management and integration, themselves derived from chromosomal repair. This genetic relationship does not compel the view that they share continuity of function.

Most apparent benefits of reciprocal recombination are rather distant from the individual genetic contexts in which they first arise (reviewed in (Felsenstein, 1974)). And, indeed, one would expect a recombining locus to be separated from any benefits it promotes a proportion of the time, at least while rare. Selection of modifiers that suppress recombination is expected, due to its contribution to the breakup of adaptive combinations, at least when considering only one side of the balance sheet.

Active recombination is a very complex process; the further apparent need for it to become widespread before significant benefit can be achieved is a substantial difficulty if recombination is to be considered a prime driver of the sexual transaction, or even if it is to become common at all within sexual (alternately haploid-diploid) species.

Active recombination occurs between homologues during Prophase I of meiosis, but the very existence of homologues and a distinct phase generating bivalents demands syngamy, a duplication step, *and* reduction, all complex operations which, must, I suggest, provide rationales of their own before we get to crossover. Despite its far-reaching consequences, reciprocal recombination is unlikely to be the reason sex (as defined) evolved.

Nonetheless, a form of recombination, and hence many of its consequences, can be easily achieved, as an unavoidable consequence of the transaction under discussion. Chromosomes segregate independently. If n – the haploid chromosome number - exceeds 1, independent fragments inevitably result, through independent segregation. The passive recombination of independently segregating chromosomes was omitted from the basic model for simplicity rather than necessity. 50% of the time, where n=2, independent segregation will cause the emerging haploids to have swapped partners. Chromosomes are not typically labelled with their parental origin. The task in reduction is simply to pair up homologues and pull them apart; all combinations resulting from that basic process are equally valid. Therefore, no gene on any chromosome has a fundamental reason to resist being separated from its previous companions in the haploid cell, nor any means to tell that this is happening. If the resulting configuration is viable, either product will do. This creates an interesting situation where the integrity of the haploid genomes that enter the diploid is not retained on exit – but this goes unnoticed by the individual evolutionary alleles. Regardless of composition, entire viable haploid sets enter; entire viable sets emerge. We still see its analogue today, despite the addition of internal crossovers that reduce the size of evolutionary fragments still further. For an evolutionary allele, one viable partnership is as good as any other.

Regardless of the order in which n>1 and cyclic syngamy occurred, the existence of both leads to a level of selection below that of the haploid genomes. Individual chromosomes are linked in a cell, whether haploid or diploid. For the duration of that linkage, they are selected to optimise or minimally damage the partnership. But on dissolution of the partnership, their only interest lies with the gene set to which they remain covalently linked – the gene set on their chromosome.

Chromosome number variation is readily achieved, through both breaks and fusions. It is improbable that, in the general case, chromosome number is under strong selection in either direction, more likely representing the drift-mediated fixation within local populations of those breakages and fusions that readily occur in the course of mechanical shepherding of these elements. Against a background of single-chromosome relatives, a broken version might be expected to have slightly reduced fitness, due to

- the requirement for each part to possess a centromere,
- unequal forces experienced during metaphase/anaphase,
- alignment issues during synapsis.

Nonetheless, even with a slight initial drag, such a break may increase in the population by drift alone. Forces acting against spread while rare act to promote it when its frequency passes 50%. As it becomes common, homozygotes will arise that

do not suffer the negative mechanical effects; heterozygosity then starts to disfavour the wild type.

Further, these break-homozygous genomes unavoidably experience reciprocal swap in 50% of reductions, providing some additional impetus by increasing the efficiency of selection without the loss of context suffered by a locus that occupies actual genetic space. A break could be treated as if it were a recombining 'gene', co-located with the site of reciprocal exchange. Like a crossover, a break will increase the variance of fitness in offspring produced by diploids homozygous for the break. But unlike an active crossover-inducing gene, which occupies a real genetic locus, a break does not risk becoming detached from the 'benefits' of its own action.

Each single-celled organism produces just two offspring at most. If we assume a genome notionally possessing two segments, denoted AB or ab, single-chromosome haploids can only produce {AB} and {ab} chromosomes, whereas a break (n=2) can produce {A,B}, {A,b}, {a,B}, {a,B}. In a diploid homozygous for the break, on the average the original haploids {*A,B*} and {*a,b*} will reduce to 50% identical output haploids, and 25% each {*A,b*} and {*a,B*}. Now if a locus on segment *a* lowers the selection coefficient of its linkage unit, its presence in the population will be diminished by selection irrespective of whether that linkage unit is {*ab*}, {*a,b*} or {*a,B*}. However, by being on a smaller unit in the latter two cases, elimination of *a* (which also amounts to positive selection of A) can be rendered more efficient, because Hill-Robertson interference, the ability of deleterious alleles to hitch-hike upon benefits provided by their linkage companions, diminishes proportionally with the decreasing size of the linkage unit.

If we add in mutation and epistasis, the general possibility exists for a build-up within a linkage unit of deleterious mutations and negative interactions by random effects. In that case, linkage unit subdivision would reduce this negative load, and increase the efficiency of this purging effect.

Under the sequential view presented here, then, we find that passive breaks and fusions can act as the 'ideal' recombining locus, reaping the benefits proposed for active loci without running into those objections demanding an adaptive explanation for their continuing persistence. If a break sweeps to fixation due to its action in demarcating a detrimental or a beneficial segment, it is then free to either drift out again or remain. A chromosome break is analogous to both recombination modifier and crossover site in an active genetic system. But, because it is essentially made of nothing, and does nothing, its mechanism needs no maintenance. Such loci can arise and disappear repeatedly while still giving many of the features provided by active crossover.

Within linkage units, Hill-Robertson and 'Muller's Ratchet' effects remain. They are two sides of the same coin. Nonetheless, the occasional passive capacity to break down the genetic complement into smaller units changes the evolutionary dynamic. Individual chromosomes are evolutionary alleles if meiosis occurs. Within this *slightly* more flexible system, we can see a further probing of lower-level selection, beyond that already provided by haploid segregation at whole-genome level. Such a population evolves *slightly* more quickly, selection is *slightly* more efficient, and adaptation *slightly* more facilitated by this passive mechanism. Given the competitors

in existence at this stage, this *slight* edge would propel the sexual clade incrementally onwards and upwards, comparatively untroubled by either asexual haploid ancestors or secondarily asexual diploid offshoots, even though it may experience some niche competition with both.

With a chromosome set, n>1, evolution has, in a passive manner, accelerated a notch, by reducing the evolutionary allele to being a subunit even of the haploid genome. This allows generation of mosaics of the best, and increase in the rate of change and the amount of standing variation a population can hold. Faster anagenesis increases the rate of cladogenesis and hence the total number of sexual taxa, some of which will interact as predators, prey or competitors. The modern sexual clade starts to emerge, in each population of which multiple adaptations can be integrated at once. Meanwhile the independent mobility of smaller genetic units allow for a better fit of subpopulations to variations in local ecology, patterning a population and rendering takeover by a 'frozen' genome in an asexual offshoot less likely still.

Even now, then, with the minimum possible sexual system worth the name, the distinctive dynamics of sexual systems start to become apparent. A significantly novel tempo of evolution has become established, simply through 'minimal sex' – cyclic haploidy and diploidy involving haploid organisms in temporary union.

The Costs Revisited

In light of the dynamics of this primitive scenario, what of the costs of sex as traditionally presented – as a penalty for remaining sexual?

Is the sexual system so far free of such costs? It would appear so. In a minimal, primitive scenario, where 'sex' consists of isogamous unions between single-chromosome haploid genomes, we can eliminate all proposed costs of sex:

Cost of Males/Anisogamy
Evidently, with no concept of division of the population into 'major' and 'minor' contributors, we can ignore this cost here.

Recombination Load
If n=1, there is no recombination.
If n>1, there is no clear reason to invoke a load. One could consider a competition between n=1 and n=2 populations. It is by no means certain that the second experiences a load when compared to the first due to breakup of adaptive combinations that it also forms.

Segregation Load
In segregating a single locus, we may break a homozygote that is fitter than the heterozygote (underdominance). We may also break positive epistasis between homologues - *trans* loci that have synergistic fitness effects. But at the same time, we break those single locus allele pairings that are less fit, and interactions of negative epistasis. As with recombination load, again through syngamy we may form the very linkages that sex is accused of breaking. It is by no means a given that the net effect

of segregation and syngamy over all segregating loci will be a genetic load relative to a particular diploid lineage that retains a given combination perennially undivided.

Cost of Meiosis

In this decomposed sexual system, we can see quite clearly that the diploid phase of the syngamy-reduction cycle is not a 'unit with interests' in an evolutionary sense – it is a temporary union of haploid genomes, regardless whether that union lasts for a second or seventy years. It does not gain such interests simply because the two genomes occupy the same nucleus. One would not regard the diploid as such a unit in intermediate scenarios such as cell-cell adhesion or two nuclei embedded in shared cytoplasm, and there is no particular reason to start doing so when all membrane barriers break down - even where this 'temporary' liaison persists through one or more diploid mitoses. The fundamental genomic units within our primitive sexual system are haploid, and they gain copies whether copied together or separately.

The view that the diploid (what we might hesitantly call the foreshadow of 'the organism') is penalised by dividing its genetic complement can be seen to be false. Similarly, individual genetic loci are not penalised by ending up in only one of the two haploid outputs. That is where they started.

There is scope for genetic conflict due to the physical proximity of the two genomes. A locus has an opportunity to disable or convert its homologue. However, as noted these mechanisms are dissipative and rare. They do not select for abandonment of reduction, which would simply swap this rather limited load for much greater ones – degradation through imprisoned transposons, Muller's Ratchet and gene conversion of deleterious recessives.

In this cost-free state, sex can establish and elaborate.

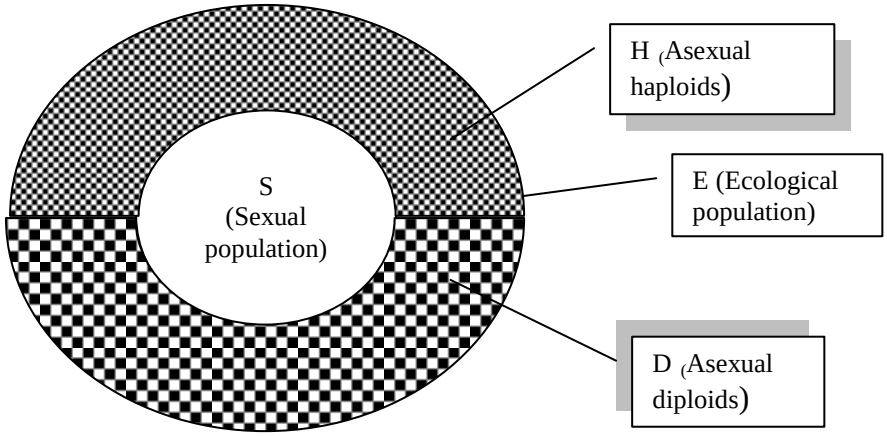
The Population and the Gene.

The mystery of sex is frequently viewed as if it were a mystery of non-adoption of perpetual asexuality. But it is important to note that if this does occur, the asexual lineage so formed *leaves* the genetic population, even while remaining part of the ecological population. This has important implications for the appropriateness of models.

In the simple haploid-diploid world I have so far described, the distinction between 'gametes' and 'organisms', and between sex and symbiosis, is still rather blurred, but we still have all the components of the life history of any conventional sexual species. The network of genomes capable of syngamy and meiosis forms a prototype population in the genetic sense: a collection of diploid individuals among which genes – evolutionary alleles – flow, albeit initially as chromosomes rather than as subchromosomal loci.

Following initiation of the primitive haploid-diploid cycle, given mitotic competence in either phase we may shortly expect to have up to three competing groups, related to the original haploid ancestor and with some degree of ecological interaction as competitors. We may denote them as H, S and D. H remains strictly haploid, S is a

derived lineage indulging cyclic haploidy/diploidy, and D is a permanently diploid derivative of S. The ecological population comprising all three we may denote as E. Representing this as a cartoon, and assuming no fundamental difference between them in ecology, we have the following set structure:



Schematic of a mixed sexual/asexual 'population' competing for the same ecological resources. Asexual individuals are represented by the black squares; sexual individuals not shown.

In practice, of course, it would be hard to distinguish sexual haploids and diploids from asexual ones, particularly given the necessary existence of haploid and/or diploid mitosis in S. Nonetheless, it remains the case that a haploid is a member of S if it retains the capacity to generate sexual diploids, otherwise it is a member of H. Likewise, a diploid is a member of S if it has the capacity to return its constituent haploids, otherwise it is a member of D.

Ecological sets too may be difficult to identify in practice, since resource utilisation is rarely discrete, homogeneous or fully congruent between different organisms. In models, however, ecological and genetic sets can and should be distinguished. Yet the interplay of genetic and ecological factors between and within sexual and asexual groups is frequently confused, a confusion that turns upon the nature of the notional alleles for sex and its absence. The same locus, having the same gene product, behaves differently depending upon the presence or absence of sex. With no reduction, the evolutionary allele is the entire genome. With sex, the allele is delimited by the extent of the independently segregating linkage unit.

Even Maynard Smith and Williams, prominent figures in the literature on both sex and on evolutionary units, are not immune to this confusion. In his textbook *Evolutionary Genetics* (Smith, 1998), under the heading "Why Not Be a Parthenogen?", he offers this model for the student:

> Consider an organism such as a herring, with equal numbers of males and females, and no parental care. In females, a gene *A* suppresses meiosis, and causes the development of diploid eggs that develop without fertilization into females genetically identical to the parent. Figure 12.5 shows that, when rare, such a gene would double in frequency in each generation. This result has been expressed by saying that there is a 'twofold cost of sex', arising from the

> needless production of males. It is clearer, however, to take a 'gene's eye view': a gene *A* that suppresses meiosis is certain to be transmitted to all the eggs produced by a female, whereas a gene *a* that permits meiosis is transmitted to only half.

I argue that this is a misapplication of his own 'selfish gene' concept, due to failure to account for a vital distinction between the two populations in respect of Williams's 'evolutionary allele'. Gene A does not spread around the population S – it cannot. It is an evolutionary allele only while independent segregation takes place. When his Gene A arises in a genome, a lineage in D is founded, reproductively isolated from S, and with all individuals reproductively isolated from each other. An ecological competition takes place, which should not be portrayed as if it were a genetic one. The validity of taking the gene's eye view at a locus stops the moment it stops being a gene in Williams's sense. The gene's eye view applies to evolutionary alleles, not to loci within them. The evolutionary allele given sex is not of the same extent as that without it, even though the genetic sequence of A is the same in both cases. If his title were expressed as: "Why Not Speciate?", the question might more readily suggest its own answer. When sex is abandoned, the result is instant speciation. Yet in his title and discussion, Maynard Smith implies that parthenogenesis can be treated as a kind of genetic switch that can be toggled on or off.

Williams took much the same approach (Williams, 1975), hoping to find general clues on the benefits of sex by examination of the 'asexual' nature of organisms such as aphids. But it is important to note that aphids are *not* asexual organisms. They simply extend the diploid mitosis of the germ line through a succession of summer bodies. They still pass through a sexual stage, and so crossover and segregation still occur periodically. Returning gametes every season makes them a sexual species, with all the genomic dynamics that this mode carries in more obviously sexual modes, in which meiosis is synchronised with production of new individuals. If an aphid lineage failed ever to return gametes, it would be a different species, with a different dynamic. The gene's eye view is an appropriate stance to take when considering how a series of mitoses might be integrated into the haploid/diploid cycle. Alleles modifiying this behaviour will be selected at that level, undergoing recombination as they do. But with permanent diploidy, the 'gene's eye view' loses meaning. If there is *nothing but* mitosis, the genetic locus is no longer an appropriate level of selection.

Selfish genes thus exist only in S. The white area represents a landscape to be colonised: the set of loci in the future sexual population. Successful colonisers can inflate their portion of the whole only within their sector, with either or both consequences of expanding the total set E or squeezing other subsets of it competitively. But they cannot get outside S and remain selfish genes in the same sense. Although we can model allele spread in all these sets as a diffusion process, the diffusing units are smaller and more mobile in S. In the dynamic, cyclic environment of S, the diploid phase provides a state that promotes the tuning of subdivisions of the genome to that phase.

Dawkins characterises selfish genes as 'levering' themselves into the next generation. They do this by getting themselves into more descendants than competing versions of their linkage unit. Metaphorically, Dawkins presents this as a desire. But that desire, however metaphorical, only extends to activities within the sexual 'bubble'; within S. The metaphorical desires of a locus within S cannot extend to the complete

abandonment of S. A gene cannot sensibly desire, even metaphorically, cessation of its own existence as a unit with desires.

Networking

An additional characteristic unique to S is due to the possibility for multiple diffusions to be in train simultaneously across the genome, dependent upon the degree of independent segregation occurring. This multi-tasking generates a process somewhat analogous to the networking and distributed processing of computing. The entire population can be seen to be 'working on' separate solutions to life's problems, the optimised results of which can be integrated into mosaic genomes, with greater efficiency than any set of lineages condemned to whole-genome, non-integrating descent. In the latter, solutions are chained together, and such composite units are evaluated and inherited as a piece through the generations. Different 'solutions' to the same or different problems may present themselves at different places, but cannot collectively find their way into the genomes of the future population. Ultimately, only one can prevail.

This is essentially a broader application of the Hill-Robertson / Fisher-Muller effect noted earlier. Any beneficial allele/detrimental context pairing whose linkage can be broken gives two alleles, one tending towards fixation and the other away from it, but the effect is not restricted to two-locus sorting. Different parts of the population, potentially subject to different selection pressures, may act upon different traits at the same time.

Even without active crossover, this networking effect will occur. Since sex, in cyclic haploidy-diploidy, inevitably reduces the size of linkage units when compared to permanent diploidy, this system has the inbuilt capacity to tune independent units at a lower level than that of entire genomes, requiring only that n exceed 1. Indeed, even where n is equal to 1, heterozygotes will contain 'better' and 'worse' versions of this chromosome. Without the sexual cycle, the latter will tend to drag the former down; with it they can be unchained.

From the perspective of the population, multiple solutions can be integrated, rendering it capable of a more rapid and 'targeted' response to selection than those lacking sex.

Variation

Concomitant with the multiple diffusions noted above, the population S is capable of sustaining more standing variation than D. In D, the variation available is constrained by the perpetual chaining of haploid genomes in pairs. Whereas population S can, for any alternative versions of the chromosome, generate both homozygotes and heterozygotes, the extent of this extra variation is dependent upon the degree of independent segregation available. More subdivision leads to a greater potential number of different genomes. This increases the resilience of S to numerous threats – from disease, from predation, from competition – when compared to D, even in the minimal system so far considered. While a given asexual individual may be perfectly

fit in sum, the likelihood is that, for at least some of its traits, the sexual population has a better version. The asexual cannot be expected to beat the sexual population in every trait, in perpetuity.

This variation also allows the sexual population to partition a heterogeneous environment more effectively, which again provides some resistance to the threat from clonal genomes.

Models that treat one or two loci at a time in a static, homogeneous contest will miss these significant features.

Evolutionary Rate

Sex is expected generally (though not universally) to increase the rate at which evolution occurs when compared to an otherwise identical asexual lineage. Smaller units in S increase the rate through both the networking and variation effects of genomic subdivision, while segregation reduces any drag from dominance effects on beneficial recessives. The greater standing variation in sexual populations provides an initial 'rapid response' buffer against environmental change. These and subsequent mutations can be more rapidly networked into mosaic genomes with reduced interference from other genes.

Random Mating

The mathematical and computational simplification of the stirred panmictic population model frequently simulates a genetically monotonous collection of individuals in a homogeneous range, differing only in the locus of interest, utilising all other resources equally and lacking hard geographical 'edges'. An individual can move in any direction in this dimensionless range and still find a mate. All potential mates are the same distance - that is, no distance - away from it.

Placing asexual offshoots into this framework – which is precisely what the classical 'twofold cost of males' treatment does - can be misleading, because the underlying vector that renders this stirring capable of assumption at all is, in fact, the haploid genome, in its quest for diploid union. In an outcrossing sexual population, particularly one with obligate diploidy and no haploid mitosis, haploids might appear to have reduced to a mere vehicle for genes. They are selected for both dispersal and mutual attraction, in both cases frequently through mechanisms operating in the diploid. Yet without sex, the stirring of a population by this means must stop. This has not prevented people from treating asexual reversion as if mating were still in place to stir the pot. If a mixture of asexual and sexual individuals were artificially stirred, a different dynamic would ensue compared to the more realistic situation in which the sexual population is stirred by dispersal and attraction while the asexual mutant, lacking these vectors, must spread out and replace the sexual population from a local nucleus driven by its own less interactive wanderings.

A real sexual population occupies a real space. The edge is constrained by an analogue of surface tension – genes at the edge cannot simply flow in any direction,

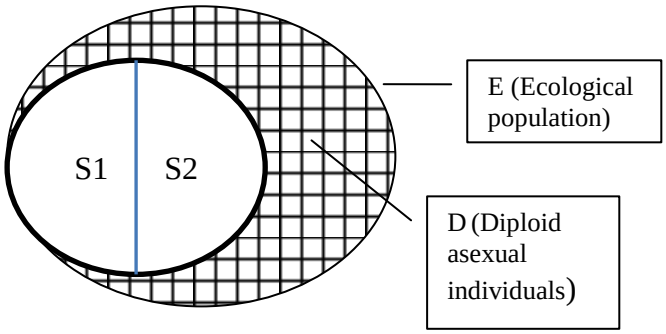
even though the individuals containing them can physically move. Crucially, they must move sideways or inwards for mates, and adaptation likewise reaches these 'edge individuals' by spread from within the range. If conditions at the edge differ from those at the heart, selection only weakly promotes alleles adaptive for edge conditions, because there are insufficient population members experiencing the edge conditions to provide for indefinite adaptation. Towards the edges, drift will be stronger than in the heart, because the available population is smaller.

Since it lacks a mating constraint, the asexual mode is better equipped to extend a range or colonise. However, it is less certainly able to supplant a resident sexual population in its home range, having no inherent advantage in exploitation of the niche, limited variation and lacking a significant vector of flow around the range provided by mating itself. Complete extinction of the parent species requires that the asexual cover its entire range, and replace it entirely. Yet because asexual individuals lack any connection either with each other or with sexual individuals, there is little that compels the ranges of the two to become co-extensive, beyond the search for food or other resources. A novel allele can spread to all corners of a sexual population's range due significantly to the vector of mating and gene flow. There is no such vector in the asexual case.

Failing to account for these geometric and vector effects in models can lead to a considerable over-estimate of the likelihood of sexual extinction. Putting both populations in a metaphorical jar and shaking it vigorously is not an accurate model of reality.

Cladogenesis

The original population S is drawn by reproductive compatibility – to be a member of S, either alone or paired, a haploid must retain the capacity to both participate in a syngamous diploid and to allow successful reduction of it. However, following a period of divergence due to some isolating mechanism that reduces gene flow, subsets of S may arise that retain the capacity for syngamy/reduction within, but not between them. This is the first speciation event: the original sexual population now forms a sexual *clade*, growing from within the ancestrally asexual haploid collection.



If we retain the artificial assumption that both species S1 and S2, and any asexual diploid offshoots D, all remain identically competitive within the same ecological

niche, then clearly any locus in any subdivision of the ecological population may mutate to cause better exploitation of that niche for its bearers. Its mode of selection is not affected by the genetic system of those bearers - but its mode of integration is, as is the rate of evolution. Only within S1, or separately within S2, can such a locus integrate into future genomes as an independent unit. These populations tend to be furnished with greater starting variation than D and tend to evolve (and hence adapt) more rapidly. As far as a gene in S1 is concerned, both S2 and D individuals are indistinguishable ecological competitors. S2, however, being a sexual species, presents a different kind of threat, because like S1 it has the capacity to build better genomes, faster, and has more standing variation. Where a superior allele exists in S1, it results in both reduction of the frequency of non-bearers in S1, and increase of bearers, potentially accompanied by an overall increase of S1. This expansion has a different effect on the two rival populations S2 and D. The latter is less well equipped to respond.

Multiple alleles are segregating in the sexual populations S1 and S2. Far fewer different locus variants are available in the clonal D populations, representing as they do one or more 'frozen' versions of a particular diploid set. At every locus in D, a superior allele somewhere in S1 or S2 may occur. Or, if it does not exist at that moment, it soon will. Even if the fittest genome in S1 or S2 becomes frozen when moving to D, it will become less fit as S1 and S2 evolve, driven not least by the ecological competitions between the subsets.

Of course, this picture is artificial. Ecological congruence was assumed to allow illustration of a 3-population dynamic with minimal complication. In practice, nothing constrains these three populations to remain closely associated with the same ecological niche – indeed, competition may itself drive divergent utilisation. The real background does not have the constant character assumed in the simple model. With a sexual ecosystem beginning to form, and ongoing anagenesis in multiple lines, the status of the threat from D, in terms of its likelihood to nip the incipient sexual clade in the bud, diminishes.

Following establishment of ecologically divergent species S1 and S2, each new D clone will of course arise from either S1 or S2. Clones may be assumed capable of production sequentially, at any time during divergence.

- D clones produced early in the divergence process will be poorly equipped to defeat S1 or S2 individuals produced later, given a period of anagenesis in both sexual lines, since these would be (on balance) better adapted.
- D clones produced late may be competitive against one of the sexual species for a time, but cannot be competitive against both, since the other has diverged.

This seems to be a difficulty for asexuality easily missed when considering simpler models. Adding this additional dynamic of a diverging sexual *clade* renders it harder for asexuals to supplant sexuality in toto. It might fight one; it cannot fight them all. This is superficially a variation on Van Valen's 'Red Queen' principle (Van Valen, 1973). In the Red Queen, species are constantly 'running to stay still'. Applied to sex, sexual species can 'run faster', thus once a sexual clade opens up, the species within it are compelled to remain sexual just to keep up with their ecological interactors.

The difficulty for asexuality amplifies as the divisions deepen, as the clade divides again and again. Gradually, a self-sustaining sexual 'bubble' opens up within the eukaryotic metapopulation. As ecological division between these descendants of the original S occurs, we may start to get a mature sexual ecosystem, where sexual predators consume sexual prey, and sexual consumers divide niches amongst themselves. Again, considerable progress can be made by sex in the minimal state so far considered, still untroubled by significant cost. Against this growing clade, replacement of sexual species on a large scale by asexual offshoots looks less and less a safe bet. Which is to say, the more realistic and complex one makes one's model, the better sex appears to do.

Asexual reversion

An implicit expectation is that if diploids arise (through syngamy or some other, unspecified mechanism), then these would somehow outcompete any sexual haploid/diploids, such that all we should be left with in Eukarya are diploids of mysterious origin. How realistic is this? It is instructive to consider the dynamics affecting such a primitively sexual species, against asexual competitors both haploid and diploid.

Haploid

Haploid reversion from the H/2H cycle would be achieved by cessation of syngamy by free-living haploids. As a threat, this is no more a difficulty than that presented by the ancestral asexual haploid lineage from which they sprang. If we are assuming a benefit to diploidy in the first instance, haploids that abandon that mode are simply back where they started, conceptually indistinguishable from the population in which sexuality arose. This need not be any threat to 'proto-sex'. The mode persists though the benefits afforded by the diploid phase, benefits inaccessible to either a primary or a secondary asexual. None of the supposed costs of sex appear to be available to favour this switch of mode.

Diploid

Diploid reversion can involve one of two types of change:

- suppression of reduction
- suppression of outcrossing.

If there is a benefit to the diploid state itself – as we must assume there must be for the cycle to initiate - the periodic coming-together of haploid genomes accesses that benefit, but there is a clear opportunity for the same benefit to be reaped by a permanent association. The cost of 'mate search' is avoided. Any 'breakup of adaptive gene combinations' is suppressed. At this stage, however, being isogamous, there is not even notionally a twofold cost of males, nor is there a twofold cost of meiosis. There is no clear advantage available to the asexual diploid here. Though it may avoid breaking up adaptive combinations, it avoids forming them too.

Further, a permanently asexual diploid encounters the following challenges:

- It always finds a close ecological competitor in residence – the diploid phases of its sexual ancestor. That competitor has
    - Variation – there are varied haploid genomes in the wider pool, and hence each diploid combination is unique. While the frozen diploid may be superior at some loci, it cannot be expected to possess the fitter variant at every locus, and beat every syngamous diploid.
    - Current adaptation – any local adaptation by S within a heterogeneous range will put an asexual D clone arising in one context at a relative disadvantage in all others.
    - Future adaptation – the sexual species has evolutionary alleles delimited by units of segregation, and hence can respond more quickly to any threat. Among those threats is the threat from a competitor, such as a derived asexual.
- It finds itself pitted against the wider ecology. In early years, this 'wider ecology' will include a small number of sexual species descended from the first. But this clade can grow in species numbers, through bifurcation and divergence. Such species may interact ecologically with any new asexual lineage, and much of the argument that applies to direct competition with the parent species applies equally to non-competitive ecological interaction with other, more distantly related sexual species. To the extent that any such diverged species are competitors, predators or prey to the frozen diploid, it will be less well equipped to deal with the contest than its sexual parent species. Both present and future variation will tend to favour the sexual species in any contest, regardless of the nature of their ecological interaction.
- It is likely to suffer from genomic degradation by gene conversion, transposition and 'Muller's Ratchet'.
- Its adaptation suffers from comparative drag due to diploid effects at a locus
- It lacks the dispersion vectors of gametes and mate search, and hence it will interfere with its own progress in a local enclave, competing with clones as well as sexuals.

Individual species may wink out of existence periodically due to asexual reversion, but once sex is in train, it seems unlikely that any mechanism could extinguish an entire multi-species sexual clade. At this isogamous stage, asexual reversion is not a significant threat, and it does not seem reasonable to suppose that this world can easily be taken over by secondarily asexual diploid lines. By the time anisogamy arrives to puzzle future diploid biologists, sex will be firmly established.

Outcrossing I

Through the assumed benefits of diploidy, a haploid indulging cyclic syngamy and reduction would increase in its overall population. As it becomes common, it will begin to encounter relatives in diploids more frequently. Genome-wide homozygosity, however, is counter-productive if the initial benefit of syngamy is related to complementation, and so there is likely to be selection favouring outcrossing – fusion with more distant, and hence more dissimilar, relatives. This principle is not without limit – genomes too distantly related are unlikely to form sensible co-operatives of

genes in the diploid state. There is thus an optimal zone; partners should be related but not too closely.

Selfing is not universally detrimental, but the circumstances in which it is adaptive tend to apply only to certain kinds of lifestyle - particularly higher plants and fungi. If – due to contingent environmental, morphological and genetic factors – selfing carries reduced risk, or outcrossing provides minimal benefit, it is a potential evolutionary strategy, although it does reduce variation and the independent character of genes due to its effect on homozygosity.

In a simple haploid syngamy-reduction system, the most local 'relatives' encountered would be those with whom one has just shared a diploid stage. If meiosis is 1-step (see below), 'selfing' would simply recreate the diploid just severed. This may seem pointless but is not particularly costly, since, ignoring gene conversion, the heterozygosity of the re-formed diploid would be the same as that prior to reduction.

However, the incidence of homozygous pairings will rise when replication steps are added, in either or both of the haploid and diploid phases. 2 step meiosis involves such a replication step leading to four haploid outputs, among which 50% of random fusions would be homozygous. It is not necessary for meiosis itself to be 2-step to have the same result as if it were. Essentially, when n, the haploid chromosome number, is 1 and there is no crossover, the net result of modern 2-step meiosis is the same as one round of diploid mitosis and two reductions, or a reduction followed by two rounds of haploid mitosis. In each case, four haploid genomes emerge from one diploid pairing, which would form homozygotes on 50% of random re-fusion events between the four.

Any pressure to avoid excess homozygosity in re-fusing gametes will act more strongly given any tendency to proliferate by mitosis *in situ* without dispersal, in either the haploid or the diploid phase – an analogue of multicellular bodies in these unicellular organisms. If mitosis produces a clonal subpopulation, this may drive selection for outcrossing of its eventual haploids. This may be simultaneously promoted by two selective forces, one repulsive and one attractive. On the one hand, cells need to disperse to avoid homozygotes; on the other, they need to do so to locate heterozygotes if complementation is beneficial. Neither of these forces need be particularly strong to achieve the result of outcrossing, since there is unlikely to be a significant opposing benefit to selfing. Response to these pressures may involve any or all of incompatibility and dispersal mechanisms and attractive signals.

Mating Type

One means by which outcrossing could be achieved in gamete populations having significant risk of homozygous pairing would be by a genetic switch enforcing complementary pairings. A clonal population of haploids would all carry the same version of a biallelic locus, but through enforcement of mating type compatibility, all diploids would be heterozygous for this locus. Asymmetric distortions from a 50-50 ratio of these subtypes, caused by differential production of each type during a haploid mitotic phase, or biased gene conversion in diploids, would be brought back

to the mean by the added premium attached to being the 'rare' allele – essentially, by the mechanism of sex ratio balance generally attributed to Fisher (Fisher R. , 1930).

Another way to achieve outcrossing is by physical dimorphism between the gametes themselves, as happens in higher plants and animals. These represent unequal distributions of cytoplasm. For the 'naked' organisms in this primitive scenario, this is probably not an option. Unicellular diploids tend to yield isogamous haploids. If they are produced by reduction of a single free-living diploid cell, any division other than 50/50 leaves the smaller at a potential selective disadvantage. A contest over cytoplasm may be envisaged, allowing a selfish allele to increase its share while selection eliminates the weaker, but any departure from the 50/50 division would soon dissipate as its increase in the population led to homozygotes, both 'equally equipped to take more than their fair share'.

Outcrossing has been noted as a cost of sex, because with mating-type partition only half the population is generally available for mating. It is not, however, clear why having a potential mate population of (say) 2,000 would be inherently better than a subpopulation of 1,000, from the perspective of any given allele. Only one partner is actually required. Further, if the reason for a heterozygous mating type is to avoid inbreeding, or to gain from hybrid complementation, that must also be included on the balance sheet. For so long as haploids can perform mitosis on their own account, they are not compelled to find a partner at all.

One- and Two-Step Meiosis.

The reduction in my model involves a 1-step meiosis, which is broadly analogous to modern Meiosis II, although involving independent homologues instead of the centromere-joined sister chromatids entering Meiosis II. The absence of a centromeric bridge is assumed not to be fatal.

1-step meiosis is almost unknown in nature, and even suspected cases are uncertain. Given that genes promoting recombination in the modern sense, as part of Meoisis I of a 2-step meiosis, are distributed universally among the eukaryote clade (both sexual and asexual members), it is probable that 2-step meiosis was present in the LCA of that clade. 1-step meiosis in modern organisms, if it happens at all, is probably a secondary derivative. If 1-step meiosis came first (as simple reduction) it was long ago superseded by 2-step.

Because meiosis has so much in common with mitosis, it is hard to avoid the conclusion that the first evolved from the second. Nonetheless, it makes a significant difference to the logic of the situation as to where in the cycle one starts. Wilkins and Holliday (Wilkins & Holliday, 2009) identify four key cell-cycle changes on the assumption that each of Meiosis I and II evolved from amendment of a complete mitosis starting with a diploid genome:

Meiosis I:

- Pairing of homologues
- Recombination between non-sisters

- Suppression of sister separation on first division

Meiosis II:

- Suppression of chromosome replication

They chose to omit one rather vital component: syngamy, without which these diploids seem to have come from nowhere, and generate haploids for no particular reason.

These five elements would be the complete requirement to derive meiosis from two sequentially arranged diploid mitoses – a significantly complex set of systems required for a single end, which the authors rightly doubt could occur together.

They discuss a simpler 'one-step' meiosis. Nonetheless, they still assume it arose from a complete mitosis, therefore requiring several steps:

- (Syngamy – not mentioned)
- Suppression of chromosome replication
- Pairing of homologues.

Wilkins and Holliday declare it hard to imagine the simultaneous evolution of the elements of even that simplified scenario, and therefore argue for the first innovation being homologue synapsis, offering selective reasons as to why this may occur prior to a 'true' reduction.

However, these speculations are only forced by a commitment to a sequence which *first* establishes diploidy, *then* results in long-term establishment of lineages of diploid cells, and *then* generates conditions which may favour reduction of homologues and syngamy. Many commentators, right back to Cleveland in 1947 (Cleveland, 1947), similarly advance the hypothesis that the diploid originated from a source other than syngamy. Due to its numerous difficulties, I am unsure why this view is so popular. As I have argued above, syngamy first with reduction shortly thereafter reduces the required novelties to precisely one: syngamy itself.

Meiosis in our scenario arises not from one or two complete mitoses in a diploid lineage (of unclear origin!) but initially from invocation of just the final step: cytokinesis. There is no synapsis; homologues may be aligned at the metaphase plate by mitotic 'rescue' pathways evolved to separate mechanically separated sister chromosomes. There is no need to suppress replication in this precursor of 'Meiosis II', since the cell treats the chromosomes as if replication had already occurred.

From this basic cycle, an additional complete mitosis would provide the basis of Meiosis I with its enhancements of replication, synapsis and homologous recombination. The original reduction process, now properly Meiosis II, is presented with paired chromatids instead of homologues. Without the addition of synapsis in Meiosis I, sister chromatids would simply align at the metaphase plate and be pulled apart – a normal diploid mitosis. By pairing up in synapsis, homologues offer a kinetochore apiece for spindle fibre attachment, one oriented to each pole, preventing the spindle fibres from attaching directly to their normal mitotic target, the single kinetochore.

But, one may legitimately ask, why the duplication step of Meiosis I?

There is something of an *impasse* in definitively explaining how the 2-step process may have become dominant prior to the LCA of the eukaryote clade. What forces if any generated the move to 2-step, and do the same or other forces continue to favour 2-step in modern organisms? Three particular features resulting from 2-step meiosis form potential candidates through which such selection may operate:

- The fact that the result gives 4 haploid outputs instead of 2
- Synapsis and recombination occur between homologues, giving a mix of reciprocal and nonreciprocal outputs.
- The replication step generates physically linked bivalents with centromeres and kinetochores, instead of the separate chromosomes available to single-step reduction.

A species that incorporated increase in the diploid phase would provide increase for both haploid genomes, exactly equivalent to their own independent mitosis. It may be that, initially, reduction had to be performed in order for mitosis to occur at all. Adding a replication step to the diploid state prolongs it, adding to the existing benefit of mutualism by allowing increase in the mutual state – an option not available to unrelated genomes such as the fungal and algal components of lichen. There is, however, no compelling reason to couple this increase to reduction as opposed to performing one or more conventional mitoses and then a 1-step meiosis. Modern multicellular organisms perform an extensive series of mitoses to increase the genomes, but still terminate the gametogenetic process with a 2 step meiosis.

No definitive answer can be provided at present as to the reason for this.

Archetti (Archetti, 2004) has advanced an argument that 1-step meiosis is susceptible to invasion by unreduced diploids.

Wilkins and Holliday (Wilkins & Holliday, 2009) favour the argument that 2-step meiosis is actually easier to achieve, although this rather depends on one's starting assumptions for the ancestral state. If meiosis originated from the cytokinetic step of mitosis, that would be easier still.

David Haig, meanwhile (Haig, 1993) offers the theory that, in combination with recombination, 2-step meiosis creates maximal uncertainty for 'selfish' genetic elements that exploit the physical proximity of homologues to distort transmission in their favour. Even in our primitive system, the equivalent of crossing over occurs, by independent segregation of multiple chromosomes.

All of these have a degree of plausibility; the reader is directed to the literature for further discussion. Nonetheless, the complexities of modern meiosis can be most easily reached if we commence with a simple cycle of syngamy and reduction.

Recombination II

Although there is no genetic element involved in multiple-chromosome recombination (see Recombination I), sex's most noteworthy feature – active crossing-over - does involve numerous 'conventional' genes that can increase or decrease their frequency in the population due to selection and drift. The number of crossovers formed can be selectively adjusted, while the siting of crossovers along a chromosome is stochastic but not equally distributed, depending partly upon the degree of local homology in a region and partly upon widely variable mechanisms of attraction and repulsion of the proteins involved in crossover initiation.

It is widely accepted that the fundamental mechanism of crossover is derived from repair pathways. In the mitotic cell, these will have been well established prior to syngamy, even in haploid lineages. Although a haploid may appear to lack a homologue, replicated sister chromosomes in fact exist through most of the cell cycle except telophase-to-G1.

Repair and crossover are both initiated by Double Strand Breaks (DSBs). In the case of crossover, these breaks are 'deliberately' induced by the protein spo11 or equivalent, a highly conserved derivative of a family of enzymes known as topoisomerases. These normally break a chromosome in order to pass a DNA strand through the gap before re-joining the ends. Spo11 lacks the latter capacities. It just opens up a DSB, whose presence causes repair-like pathways to initiate. In non-meiotic cells, DSBs tend to be repaired from the replicated sister (where available) or by the potentially mutagenic NHEJ pathway, rather than using the homologue which will (in a slight terminological confusion) have less sequence homology than a replicated sister. By contrast, in meiosis the preferred pathway is homologous repair using the homologous (not sister) chromosome. This choice is central to all the far-reaching consequences of recombination, because it is the cause of reciprocal swap between segments of parental chromosomes.

Given the profound consequences of this, it is tempting to see those consequences as the reason for the swap. But in fact the reasons may be more mundane, and the consequences incidental. When homologous repair is performed, the two chromosomes become joined in either a single or a double Holliday Junction (Holliday, 1964). This is a mobile structure joining two double strands of DNA such that one strand of each chromosome pairs with its complementary sequence on the other. The structure can be resolved back into separate chromosomes by two judicious snips. There are two options for these snips with the same result at a given junction.

If a single Holliday junction is formed, the result of resolution is non-crossover (though some gene conversion occurs). If two junctions are formed, half of the products of resolution are crossover and half non-crossover. This is due to the polarity of the resolution of each junction in the pair. There are four possible resolutions in total, two at each site. If we call one '+' and the other '-', crossover occurs with +/- or -/+, but not with ++ or --.

The resolution process cannot actually tell which product it is going to create. There is no information upstream or downstream of a junction that tells the process which chromosome is which. Both crossover and noncrossover products are functionally identical. There is no requirement for any gene to continue to be linked to a given set, provided that the new set works. Therefore, even if reciprocal swap were a

requirement, the only way to achieve it is to initiate twice as many DSBs as are needed for crossover. It is interesting to compare this to the situation in both passive recombination in multi-chromosome genomes, and in crossovers formed during resolution of homologous repair using sisters. It is tempting to think that something should 'care' which way things go, for long term reasons, but the information to make a decision is unavailable. It may in fact be immaterial whether products are recombinant or not, despite far reaching consequences.

Having described the basic process, let us turn to its rationale. Why does crossover occur at all, and why does it prefer the homologue, which is normally last on the list of candidates for DSB resolution outside of meiosis?

There are three main classes of theory for the presence in populations of genes involved in active recombination involving crossover (they are not mutually exclusive):

- Cytological role
- Population effects
- Genomic conflict

Cytology

The preference of the crossover of meiosis for the homologue, rather than using NHEJ, or HR with a sister template, creates the strong impression in many minds that this must be what crossover (and often, the whole of sex) is 'for' – creating recombinant haploids from homologues. But it is certainly not the objective of the ancestral repair process, having only sisters where 'to-swap-or-not-to-swap' is not even an issue, both products being indistinguishable. I would argue that the same is true for most loci swapped in a reciprocal exchange between non-sisters. The upstream and downstream companions of a gene are of no interest to it if there is equal likelihood of their replacement by a viable gene set under crossover. Further, most genes are incapable of influencing matters anyway.

In fact, crossovers *must* be between non-sisters if they are to play a cytological role in segregation of tetrads and/or homologues. The replicated sisters are each connected by a centromere, which provides tension for segregation in Meioisis II but has the opposite role in Meiosis I, ensuring physical passage of the bivalent sisters to the same spindle. The centromeres are offset from the metaphase plate in Meiosis I and so something other than a centromere must act as a centre of tension between the homologues. That role is taken by a physical crossover. Reciprocal swap between the homologues is an inevitable side effect.

Recombination also provides a possible mechanism for *identifying* homologues. There must be some means of 'assessing' the sequence of chromosomes in order to pair them up, and inserting a few judicious DSBs and allowing the repair pathways to cause interaction is one possible means to do that.

In the modern cell, we have complex synapsis, which appears based upon less crude mechanisms of homology search than the rather clumsy and mutagenic method suggested. Some modern meioses are fully achiasmate, though organisms that do

without crossover tend to do so in one sex only, for example *Drosophila* males and *Bombyx mori* females. Other mechanisms substitute for chiasmata in ensuring disjunction – the synaptonemal complex in *Bombyx*, chromosome 'territories' in *Drosophila*. These are derived states, and it is probable that the ancestral state was chiasmate.

Population effects.

The consequences of crossover are dramatic, once it is established in a population. Introducing crossover into genetic algorithms greatly increases their search speed and exploratory capability, and it seems unlikely that the inspiration for this method has no similar consequences in nature, stretched metaphors notwithstanding. Pathways that cannot by explored (or are more slowly explored) by stepwise mutation can be circumvented by combining genomes that have been through independent selection. Standing variation at each locus increases, as does the variation along the linkage unit, due to the independent segregation introduced by this fine slicing. Repeat recombination into different genetic backgrounds exposes an allele's 'true' selection coefficient, independent of Hill-Robertson effects, allowing genomes to be purged of their detriment and increased in their benefit. This also unchains the 'selfish gene'. The evolutionary allele drops another level, having first gone from haploid genome to chromosome, it now drops to coincide more closely with the molecular biologists' and geneticists' genes.

However, the nature of these population benefits is such that they cannot be supported as a *primary* driver of recombination. Recombination must become common by some other means, and the demands of cytology provide a strong contender for the selective force responsible.

Regardless of selection for the basic mechanism of recombination, there is a role for modifiers, and crossover placement is under extensive control. The issue facing a genome is how to establish, with the genetic material presented, sufficient DSBs that crossover occurs, but not so many that it becomes self-defeating and mutagenic. Each chromosome pairing is unique, with a highly variable amount of homology between the partners, and a variable presence of DSB-inducing 'hotspots' and other circumstantial variables which play a part in establishing the initial DSBs. A certain amount of 'experimentation' goes on before the cell commits itself and crossover occurs.

Population effects undoubtedly affect the presence of modifiers, but few of the crossover distribution patterns can be clearly linked to variance effects or linkage equilibrium caused by reciprocal swap itself, as opposed to circumstantial constraint or selection for cytological effect. Crossover seems to have the status of a 'necessary evil' at the individual level:

- The presence of a crossover tends to deter others from forming nearby – their distribution with respect to each other does not accord with a random expectation. This is consistent with a cytological role, which requires a non-random distribution. A fully stochastic process would place some crossovers too close together, and leave some chromosomes untouched while doubling up in others. But in fact, we get a fairly consistent 1 per chromosome.

- Crossovers are usually rare in centromeric regions, and frequently elevated towards the telomeres. This is probably due to their greater mechanical usefulness towards the ends of the arms, away from the centromeric site of spindle attachment. One may also envisage orientation and spindle-targeting difficulties arising from placing the centromeres of two bivalents too close together on the metaphase plate, which would occur given pericentromeric crossovers.
- Crossover rates vary between the sexes in many species. This may reflect differentials in the structure of meiosis, or may indicate an influence of mitochondria or sex-linked traits, but is unlikely to be related to any population effect since chromosomes spend half their time in each sex.
- Crossover rates are higher where overall homology is greater – for example, in domestic animals. Maynard Smith attributes this to hitch-hiking under strong directional selection (Smith, 1998), though this implicitly demands a rather unlikely situation where at least one active modifier exists per chromosome, since all chromosomes show this effect. More likely this is a circumstantial effect, since homology influences the siting of DSBs.

Genomic conflict

Meiotic recombination interferes with the operation of certain 'selfish' elements which proliferate by distorting Mendelian transmission in their favour. Because many mechanisms require identification of 'self' and 'non-self', meiosis offers uncertainty about futures – particularly when performed in the '2-step' version, where four different haploid outputs are formed. 'Killers' frequently hide near centromeres, for example. However, these ingenious ideas are unlikely to offer a strong selective advantage. The mechanisms dissipate anyway in time, because successful versions tend to spread and meet more and more copies of themselves, and fewer and fewer examples of 'non-self'.

The Role of Mitosis.

Syngamy and 1-step reduction do not provide any increase, although with a 2-step meiosis, the initial step provides increase – 4 haploid outputs are produced from 2 diploid inputs. This, however, relates to the dynamics of the process rather than contributing significantly to increase. The principal mechanism available for increase of a genome remains mitosis, which in ideal conditions doubles the population every generation. Such increase in the diploid state may not originally have been available, although there is no particular consequence for theory either way. Whenever it did become available, three options then presented themselves for increase in the syngamising haploid population:

Increase in the haploid phase alone
Increase in both haploid and diploid phases
Increase in the diploid phase alone

In this model, these form a probable evolutionary series from early days to the familiar pattern in many modern species, both multi- and unicellular. Haploid mitosis is comparatively rare in eukaryotes, among those both single-celled and multicellular (with the notable exception of the fungi and 'lower' plants). Once increase in the diploid phase became established, haploid increase – indeed, most of haploid existence – seems to have become increasingly unnecessary. The haploid shrank to a barely-noticeable transitional phase, but there is compelling reason for the diploid to become permanent. Diploid mitosis can form part of the strategies available to the haploid genomes without any associated need for them to cease being produced at all.

If diploidy started out 'for' the joint benefit of haploid genomes, this will remain the case even if diploid pairing becomes by far the dominant stage. The haploid genomes simply share the benefit in harness. They have no reason to take the step of staying bound together eternally with all the penalties previously noted of doing so. Diploidy, for a sexual lineage, is always temporary, by definition. Though permanence may happen from time to time, it does not seem reasonable to assume it should be widespread, and force sex to 'pay its way' against some illusory costs.

Multicellularity

Although there seems little rationale for a permanent adoption of diploidy, a significant use for this mode is found in the switch to multicellularity, which has occurred at least 16 times independently.

In a simple alternation where the diploid and haploid phases are free-living cells, every such diploid can return haploid gametes. However, if there is adhesion of diploids, in a colonial form, an interesting new dynamic arises, whose coherence is ensured by sex. The adhered diploids may make an additional contribution to fitness beyond that available to free-living lines. They may, in forming a larger entity, be less susceptible to predation, less susceptible to invasion by disease, better equipped to garner nutrition, and be capable of specialisation by differentiation into different cell types using the same genome(s).

If the organism is asexual, there is no particular rationale for the agglomeration of such a unit. Each clonal cell has no interest in other cells that share its genome. There is no greater advantage to helping a clonal relative to reproduce when compared to reproducing oneself. If, however, there is a haploid reduction, a given diploid gene set can either reproduce on its own account, or assist cells that ultimately release its haploid constituents by proxy.

When the time comes to release haploids, in a primitive system all cells may be totipotent. Yet there is no need that this be so. The cells in this primitive colony are all clonally derived from the zygote – the initially fused pair of haploid genomes. Every cell lineage, be it somatic or germ line, must retain the capacity for mitosis as part of its role in developing specialised tissues. But there is no need for all to retain the capacity for meiosis. The release of gametes can *itself* become a specialised function.

If one considers the case of a population in steady state, two net haploid inputs can only get a maximum of two net haploid outputs to maturity. It does not require every

cell line to reduce in order to achieve this. If a colony of 16 diploid cells produced 32 haploid outputs, these would be culled to 2. If, on the other hand, the generality of diploid cells surrenders its capacity for meiosis, and specialise for other roles, those net 2 gametes can be produced by specialised tissues, freeing the remainder to differentiate to other roles.

The diploid lineage may become a *soma,* wrapping itself round and nurturing the germ cells. The germ cells are free to specialise as germ cells, likewise those cells that forego their direct reproduction. It does not matter whether a diploid reproduces on its own account, or assists a germ cell to reproduce the same genome, in a haploid-producing system the result for the constituent gene copies is the same.

This is a case of Hamilton's kin selection rule rB > C. Since the relatedness r of any diploid cell to a diploid germ cell is 1, it requires only that benefit B exceed cost C. A cell does not have to reproduce itself. However, this cohesion is cemented by the haploid output. The stability of the system does not simply derive from relatedness; if that were all that was required, prokaryote sisters would assist each other. It is instead relatedness in combination with a specialised mode of output. Diploid cells can all do mitosis; they cannot all do meiosis while at the same time specialising for a somatic role.

In the above scenario, the choice for a diploid somatic cell was to replicate itself or leave that to cells in specialised germ tissue. The organism – the multicellular diploid phase – forms part of a cyclic system, whereby haploids fuse, perform diploid mitoses generating both somatic and germ cells, and then reduce to haploids via the latter, which disperse to recommence the cycle. The function of this multicellular mode, is ultimately to generate haploid genomes in number. Multiplication is provided by mitosis in the germ line, protection and differentiation by mitosis in the soma. The ultimate product of this cyclically colonial process is haploid outputs in bulk.

In a purely diploid line, conversely, there is no haploid reduction. All cells in the colony are 100% related to each other, but none has a reason to promote a relative's reproduction in place of itself. When reproduction is the function of specialised germ line cells, other diploids no longer need to maintain reproductive competency, even while they must remain mitotically competent. But without specialised cells performing gametogenesis, a purely diploid colony has a reduced capacity to specialise, diminished by the need to remain capable of producing offspring. A colony composed solely of diploid cells would lack the cohesion brought by the gametic exit.

Of course such colonies are frequently produced, as asexual offshoots of sexual lineages. But these colonies have been generated and tuned by sex; they did not become colonial without it. When they subsequently abandon it, they have stopped doing the very thing which led to their cohesion in the first place. They don't immediately disintegrate, but the many specialisations of sex sit ill with their new-found mode, and they are ill-equipped to divest themselves of the these adaptations to a sexual life cycle.

Since the principal anticipated threat is from multicellular diploids, it is curious to note that these could not even exist without sex, the very thing whose existence they are supposed to render puzzling.

The Rotifer

> *If I were ever to essay a serious assault on the 'essential feature of the situation that is being overlooked'*[2]*, it is hereabouts that I would start. And I would listen again and again to the Rotifer's Tale. These [...] may hold the key to the outstanding paradox of evolution. What's wrong with sexual reproduction, if the bdelloid rotifers have run with it for so long? Or, if it's right for them, why don't the rest of us do it and save the massive twofold cost of sex?* Richard Dawkins, The Ancestor's Tale

Bdelloid rotifers are frequently cited as a kind of litmus test for theories of sex. Because, whatever one may say about the tendency of asexual lineages to fizzle out on an evolutionary timescale, bdelloid rotifers (and other groups such as oribatid mites and darwinulid ostracods) buck this trend by having survived quite happily for tens of millions of years without, so far as can be determined, any meiosis at all.

On the view presented here, it is not entirely clear why such arcane groups should trouble us much, but given the status of these organisms in the debate, we cannot just brush them aside. There must be a reason why asexual groups tend not to last long, and why these do not follow that trend.

Nonetheless, it is important to note that none of these groups is primitively asexual. All have sexual ancestors and relatives. As a reason to ponder the very existence of sex, organisms that cannot even come to exist without it seem a somewhat paradoxical choice. We seem to be being invited to consider only the two horns of a dichotomy: either sex is universally beneficial, in which case there should be only sexual organisms, or (absent a twofold benefit) it should be universally detrimental, in which case there should only be asexual descendants of anisogamous sexual organisms plus sexual isogametes. That seems an unjustifiable expectation. What precludes those intermediate situations in which other relative proportions of species exist? Such as the one we have, in which a tiny handful of species represents a counterexample.

The general problems for secondarily asexual lines comprise

- Gene conversion
  - increase in deleterious homozygosity
  - loss of complementation
- Slower rate of evolution
- Interference of loci
- Muller's Ratchet
- Selfish genetic elements
- The resident sexual
- Coevolution in the wider, frequently sexual, ecology.
- Heterogenous niches

[2] A reference to John Maynard Smith's dry observation in a paper: "One is left with the feeling that some essential feature of the situation is being overlooked" (Maynard Smith, 1976).

These are, I think, perfectly adequate to explain the general trend towards elimination of asexual lines on an evolutionary timescale. Even asexual offshoots of a dioecious species suffer these penalties. It is not at all certain that producing twice as many grandchildren is sufficient to cause the extinction of the parent, from a standing start, in every single case, nor even in most of them. To the extent that this happened, we would not see it. If the tendency is for asexual lines to go extinct, this may happen before or after extinction of the parent population, if they compete. In both cases, the modern eukaryote clade would be dominated by species to which this had not happened, even if 'the massive twofold cost of sex' were a real challenge.

But what about the rotifers?

A near-universal issue for asexual lineages is gene conversion. This is a consequence of homologous recombination during repair. It causes loci to become increasingly homozygous, which can expose deleterious recessives and decrease the complementation of heterozygous alleles. If chromosomes are too closely related, this results in reduction of 'vigour', or inbreeding depression when there actually is breeding. Gene conversion causes a kind of inbreeding depression in a non-breeding setting. Rotifers, however, appear to lack recombination between homologues. The haploid chromosome sets have become highly divergent. It is difficult for us to identify their homologous chromosomes; it would be equally difficult for the repair system to locate homology for repair. Presumably repair uses sister chromosomes outside of the G1 phase, in much the same way as prokaryotes, but only end joining within G1.

It's not clear whether the lack of HR is a consequence of the divergence of chromosomes, or whether its suppression was a preliminary to it, but either way, rotifers have a greater than expected amount of sequence divergence (Welch & Meselson, 2000).

Many of the other challenges in the rotifers are dealt with by their indulging a form of 'parasex' through gene transfer. Evidence for extensive gene transfer has been found (Deborteli, et al., 2016). The mechanism is currently unknown, but is likely to be a significant cause of the sequence divergence noted above. If sequences diverge too far, homologous repair may become positively mutagenic, and not merely a source of increasing homozygosity.

And finally, we can consider their ecology. Rotifers tend to blow around the world in dust, colonising freshwater puddles until these evaporate in their turn. They do not find a population of resident, ecologically indistinguishable sexual organisms, over whose continued existence we might wish to puzzle. It is not clear whether the rotifers' sexual immediate ancestors have diverged or have gone extinct, but in either case, the rotifer's life cycle and ecology would suggest that they would not be expected to be in close competition with sexual forms globally.

A related group, the monogonont rotifers, may represent the situation in the immediate ancestors of the bdelloids. They have a series of parthenogenetic generations, but intermittently return haploids. Occasionally, a pathenogenetic lineage becomes permanent, but this has not reached the widely divergent state we see in the

bdelloids. If we wish to ponder the bdelloids, it seems more appropriate to consider the monogononts than to export their peculiarities to the whole of the sexual world.

Taking these considerations in sum, rotifers do not seem particularly troubling for the views expressed here nor, *pace* Dawkins, do they offer much insight into the reasons the bulk of Eukarya are not like them. The problem is not that successful asexuals exist, but the idea that there is a 'massive' twofold cost of sex.

Gender

Sex in multicellular plants and animals is almost universally characterised by two genders. These may be either on the same (hermaphrodity) or different (dioecy) individuals. Gender fundamentally relates to the fact that haploid gametes produced by female or male parts are of relatively larger and smaller size respectively – sperm or pollen are produced by male gametogenesis, egg cells by female. These perform different, but naturally complementary, roles.

- The smaller male gametes can be produced much more cheaply per cell, and hence in great numbers. Because they must fuse with an egg, it is the number of the latter that limits the production of zygotes, but production of large numbers of cheap gametes is fuelled by competition between males for that limiting resource, and by the need for insurance against losses.
- Being smaller, they can take on the role of 'seeker' – if the larger cells stays put, while the smaller is motile or readily dispersed by the wind or currents, the chance of a successful encounter is enhanced for both cells.
- The larger egg cells tend to be packed with the initial nutrition that provides for the early mitotic divisions of the zygote as it begins to build the soma of the next generation.
- Anisogamy assists outcrossing, due partly to the tendency to dispersal of the male gametes and (when associated with dioecy) partly to the inability of individuals in such a population to self-fertilise.
- Additionally, there is a general tendency among gendered organisms for mitochondria to be passed through the female line only. This is not merely attributable to their greater size. Many sperm are furnished with mitochondria for respiratory competence; these are actively eliminated on generation of the zygote. Similar mechanisms exist in unicellular organisms, which do not exhibit extremes of anisogamy. Nonetheless, given selection for reduction of mitochondrial competition by squeezing through one line only, the larger female gametes are the vehicle of choice.

It is clear that anisogamy is an adaptation that is only available to multicellular organisms. Free living cells cannot avail themselves of either the large-scale generation of dispersible male gametes, or the nutritional environment of female egg cells. Non-equational division (unicellular hermaphrodity) would lead to a single, small male gamete, easily lost to predation. Gender specialisation (unicellular dioecy) is equally hard to justify: a species that produces a pair of motile male gametes from one diploid cell, and a pair of large female cells from the other type, lacks any rationale for the transition to this state from isogamy. The smaller male cells are still too few to compensate for losses, and vulnerable; the larger female cells are costly.

Nonetheless, it is useful to imagine such a simple system for the purposes of illustration of what appears to be a fundamental flaw in the assumption of a twofold cost of anisogamy. Suppose we did have a unicellular species that discovered a beneficial means of asymmetric partition. The ancestral isogamous species is free of the twofold cost, on arguments presented earlier. Suppose that the initial asymmetry is a 49-51% split. Can we really suppose that this transitional species has suddenly started suffering a twofold cost? It is almost isogamous. The same argument can be applied throughout a theoretical transition through increasingly asymmetric divisions. There is nowhere that the twofold cost becomes apparent, if we take this gradual approach. And yet we are assured that a significantly asymmetric case - say 1%/99%, - would suffer from a twofold cost of anisogamy. This diploid cell splits its genome equally between the two haploids, but invests most of the cost in the larger. It would save little by avoiding production of the smaller gamete.

We could not realistically expect this precise scenario to occur in nature. Such anisogamy cannot be sustained in a unicellular species. Furnished with a multicellular soma, on the other hand, we have an environment in which a move to anisogamy can be sustained. We have the possibility of specialisation into different germline organs. Mitosis can be used to produce male gametes in great numbers and little cost (relative to the cost of the soma itself). The fewer female gametes, meanwhile, can be furnished with nutrition from the soma.

Again, however, it is not at all clear that a gradual transition from isogamy to anisogamy would be accompanied by the incursion of a twofold cost, either gradually or sharply at some tipping point. Relative gamete size *per se* does not cause this issue. The twofold cost is not a cost of anisogamy. It can only arise when the population is divided into male-gamete and female-gamete individuals, as a cost of – if anything - dioecy.

Dioecy/Hermaphrodity

Given the association with multicellularity, a number of different strategies are available for the production of the male and female halves of the sexual equation in an anisogamous species. The extremes are represented by hermaphrodity and dioecy. Hermaphroditic species have the capacity to generate both male and female gametes in the same diploid individual; dioecious species have separate individuals for each sex. Intermediate situations can exist, where some individuals are hermaphrodite and others of one sex only.

Hermaphrodity is the rule among plants. Only about 4-5% of outcrossing plant species are properly dioecious; the remainder produce either all hermaphrodite individuals, or a combination of hermaphrodite and single-sex individuals.

The pattern is reversed in animals: about 7% of species are hermaphrodite, mainly among the invertebrates. Those vertebrates that do display hermaphrodity are sequentially hermaphroditic: they do not produce both kinds of gamete at the same time.

Self-fertilisation is a significant hazard of hermaphrodity. It reduces the variability of progeny. Furthermore, where there is independent segregation of loci in meiosis, it increases the chance of expression of deleterious recessives, being the ultimate in inbreeding. Therefore, numerous mechanisms promote outcrossing, whether through direct selection against selfing or as an incidental consequence of a different adaptive constraint. Sequential hermaphrodity, timing differentials of gamete production and self-incompatibility mechanisms all serve to reduce selfing. Dioecy can be added to this list. Even if not the prime cause of this mode, the enforcement of self-incompatibility is at least an inevitable secondary benefit.

There is a simple transition available from hermaphrodity to dioecy, and plausible intermediates are represented by many plants and by the model nematode *C. Elegans*, where single-sex individuals occur and can mate successfully with the more usual hermaphrodite individuals in the population. The transition to full dioecy would require a second single-sex lineage of the opposite orientation and the replacement of ancestrally hermaphroditic lines by these two single-sex forms. Such mixtures are rather rare, although that may reflect a rapid replacement of hermaphrodite individuals when dioecy does arise in a species.

The curious thing about the evolutionary succession being presented is that only now, near the very end of the discussion, do we need to worry about Maynard Smith's 'twofold cost of males'. It can hardly be a central 'mystery of sex', which has been free thus far to build a substantial varied, divergent clade of isogamous unicellular organisms, a great many partially or fully hermaphroditic plants, and all hermaphroditic animals. The dioecious few seem rather exceptional, not a reason to expect sex to suddenly start paying double, globally. If we can account for sex in 99% of the Eukaryote Tree, fundamental as it was to its construction and evolutionary persistence, the existence of dioecy in the part of the tree that interests us most seems simply anomalous, not a reason to query the entire enterprise. It is an illusion of perspective.

Even taken head-on, there is no compelling case for a universal expectation of asexuality eliminating dioecy wherever it arises. An asexual 'female' lineage arising as an offshoot of a dioecious species has much the same difficulties as one arising in an isogamous population. It finds a resident sexual, locally adapted throughout its range, with a pool of variation, whose genes are tuned to the demands of both male and female bodies. In fact, those somas are strongly tuned towards generating gametes and locating their complement, and their very coherence depends upon sex. Asexual offshoots are likely to retain many sexual adaptations, redundantly, and suffer all the difficulties previously noted for asexuals derived from isogamous species.

The theoretical twofold cost of *males* in an artificial model should not lead us to conclude that *sex* is mysterious. This exposes the peculiar paradox at the heart of much 'mystery of sex' thinking. We are to suppose that sex is a puzzle because, whenever it becomes established, its asexual descendants should always wipe it out. Yet the threat most frequently referenced - multicellular asexual 'female' lineages in a dioecious species – cannot even form without sex having become well established. Proponents of this view implicitly argue not that sex should never have existed, but that it should have existed but then been destroyed by its children, leaving an asexual multicellular biosphere, inexplicably diploid.

Offspring

When a sexual multicellular organism reproduces, it produces haploid gametes, which fuse with those of another individual to produce a diploid zygote. It seems that each individual has halved its genetic contribution through biparental inheritance. However, the average replacement for these two individuals is two offspring, not one. If the population is in steady state, they are no more 'halving' their genes than a unicellular diploid is when producing both outputs simultaneously. Two haploids emerge from isogamous division, but equally a *net* two haploids emerge from offspring production, even if those offspring are produced at different times. Some genes are in neither, some are in both, but averaged over the long run, the expected outcome is to be in half of them. A gene on a haploid locus has swapped certainty of being in one haploid or the other for a 50% chance of being in a given gamete. Averaged over multiple lives, no allele would suffer a penalty - even when half of those lives are male.

Symmetry

The fundamental stability of sex derives from its central symmetry. That symmetry is not immediately apparent in complex modern systems, but primitive, isogamous sex can be seen to be more clearly symmetrical. Sex is an alternation between haploid and diploid phases, a temporary union of two haploid genomes in one cell. The centrall symmetry between the partners is stable because any contests that arise are between partners of equal strength. Neither can gain the upper hand for more than a few generations before symmetry pulls them back towards the centre ground.

Other, more readily sustained asymmetries, have appeared between

- Diploid and haploid phase length
- Gender, at gamete and macroscopic levels
- Parental investment
- Simultaneous vs serial offspring production
- Genetic conflict

But through all of the modern complexities, at the heart remains that cycle of symmetric union and dissolution. None of the asymmetries forces a conclusion that the alternation must ultimately be replaced by one component of it repeated in perpetuity.

Final Remarks

This argument is perhaps simply a lengthy expression of puzzlement. I don't feel I have brought any especially novel arguments to bear; I am simply puzzled as to why the seemingly obvious perspective shift from diploid to haploid has not been addressed before or, if it has, why it has not become currency in the biological world. Barely a paper goes by without reference, in its introduction, to the twofold cost of

sex and the expectation that a compensatory twofold benefit must be located. And yet, as I have shown, the twofold cost of meiosis is defeated right at inception by the well-known absence of such a cost in isogamy. As such, a sexual clade is free to build largely untroubled by theoretical threats of such magnitude from asexual reversion. While sub-diploid genetic units are produced, they function as 'selfish genes' whose extent is determined by the degree of independent segregation, integrating population-wide solutions into mosaic diploid genomes. Locus-based models that ignore the role of recombination in the very nature of evolutionary alleles, and that implicitly stir the population efficiently even in the absence of sex-associated vectors, give a false picture of the genetic costs.

Secondary asexuals are a different species. The 'selfish locus' within a sexual population loses its identity when it causes reversion, and the competition becomes solely ecological and interspecific. Such asexuals as occur are not 'stirred into' the population: they have abandoned mating, and therefore have lost an important vector that underlies the very assumption of classically panmictic populations. They must colonise starting from a restricted area into a wider range already occupied by variant resident sexuals and a wider ecology dominated by sexual species. They may successfully colonise outside of the range of the resident sexual, but displacement of the latter is a much harder nut to crack. It would be far easier if they really were stirred into the population as the models have it, just as metastasis renders it far more likely for an individual to succumb to cancer.

Costs of recombination and segregation are again viewed from an erroneous perspective. If an advantageous combination occurs in the diploid phase and is favoured by natural selection, the ability to originate such combinations is a *benefit* of sex, not a cost, granted that the same process occasionally acts against itself in the breaking up of its own adaptive combinations. To the extent that selection causes increase of alleles involved in positive epistases and decrease in negative, there will be a corresponding increase or decrease in the likelihood of re-assembly from the gamete pool. And hence, taking consequences as a whole, the 'loads' attributed to these processes appear to be based upon incomplete accounting. No net cost can be reaped by abandonment of reduction to gametes in the general case, when one considers a realistic repertoire of allele interactions. This statement is of course difficult to support in other than a textual manner because of the difficulties of generating a 'realistic' genome in a model scenario.

Finally, the cost of males, the source of much puzzlement for Maynard Smith, can be seen to be a strictly minority case. Most plants and many animals do not even have males as separate individuals. Where they do, they pay their way to some extent as vectors of dispersal. And, such species have all the resilience of resident sexuals in isogamous contests. There is no genetic asymmetry even in dioecious species; as far as genes are concerned, isogamous and anisogamous sex 'look the same'.

Even the simplest haploid-diploid cycle imaginable has all the fundamental characteristics of a modern sexual system. It generates locus-level evolutionary units, diploid individuals, recombining populations, and a new mode of cladogenesis where such populations split through mutual isolation. The sexual eukaryote clade rumbles into life, characterised by symmetry and internal genetic stability. Eventually, the purely haploid ancestors are gone, only the cyclically diploid ones remain, punctuated

by occasional secondary reversions subject to genetic degradation and the ecological difficulties presented by both resident relatives and the wider ecology. Once it gets going, sex is self-sustaining and stable.

Asymmetries arise, between the generations increasing the relative persistence of the phases, between the relative role of mitosis in each phase, and between the genders where these exist. Crossover becomes established, for local cytological reasons but with far-reaching consequences. Haploids become diminished, sliced and diced, and we diploids wonder why we bother with them at all. Meantime, the ancestral haploids continue to make us binary organisms, and then file obediently into gametes in the germ line, a transaction that has not changed in its fundamentals since its inception.

Bibliography

Archetti, M. (2004). Loss of complementation and the logic of two-step meiosis. *Journal of Evolutionary Biology, 17*, 1098–1105.

Cleveland, L. R. (1947). The origin and evolution of meiosis. *Science, 105*, 287-289.

Dawkins, R. (1976). *The Selfish Gene.*

Dawkins, R. (1982). *The Extended Phenotype.* Oxford: Oxford University Press.

Deborteli, N. e. (2016). Genetic Exchange among Bdelloid Rotifers Is More Likely Due to Horizontal Gene Transfer Than to Meiotic Sex. *Current Biology*, 723-732.

Dolgin, E., & Charlesworth, B. (2006). The Fate of Transposable Elements in Asexual Populations. *Genetics*, 817-827.

Felsenstein, J. (1974). The Evolutionary Advantage of Recombination. *Genetics*, 737-756.

Fisher, R. (1930). *The genetical theory of natural selection.* Oxford: The Clarendon Press.

Fisher, R. A. (1928). The Possible Modification of the Response of the Wild Type to Recurrent Mutations. *The American Naturalist, 62*(679), 115-126.

Haig, D. (1993). Alternatives to meiosis: the unusual genetics of red algae, microsporidia, and others. *Journal of Theoretical Biology*, 15-31.

Hamilton, W. D. (1998). *Narrow Roads of Gene Land Vol 2: Evolution of Sex.* Oxford University Press.

Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genetics Research*, 282-304.

Kondrashov, A. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature, 336*, 435 - 440.

Muller, H. J. (1932). Some genetic aspects of sex. *The American Naturalist, 66*, 118-38.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research, 1*, 2-9.

Smith, J. M. (1978). *The Evolution of Sex.* Cambridge University Press.

Smith, J. M. (1998). *Evolutionary Genetics* (2nd ed.). Oxford University Press.

Speijera, D., Lukešb, J., & Eliáš, M. (2015). Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. *PNAS, 112*(29), 8827-8834.

Tucker, A., Ackerman, M. S., Eads, B., Sen Xu, S., & Lynch, M. (2013, Aug). Population-genomic insights into the evolutionary origin and fate of obligately asexual Daphnia pulex. *PNAS, 110*(39), 15740-15745.

Van Valen, L. (1973). A New Evolutionary Law. *Evolutionary Theory*, 1-30.

Weissmann, A. (1891). *Essays upon heredity and kindred biological problems.*

Welch, M. B., & Meselson, M. (2000, May). Evidence for the Evolution of Bdelloid Rotifers Without Sexual Reproduction or Genetic Exchange. *Science, 288*(5469), 1211-1215.

Wilkins, A. S., & Holliday, R. (2009, Jan). The Evolution of Meiosis From Mitosis. *Genetics, 181*(1), 3-12.

Williams, G. C. (1966). *Adaptation and Natural Selection.* Princeton, N.J: Princeton University Press.

Williams, G. C. (1975). *Sex and Evolution.* Princeton, N.J.: Princeton University Press.

Wright, S. (1929). Fisher's Theory of Dominance. *The American Naturalist*, 274-279.
Wynne-Edwards, V. C. (1962). *Animal Dispersion in Relation to Social Behaviour*. Oliver & Boyd.