Imperial College London

Evolution of a Selfish Genetic Element: The 2 Micron Plasmid of *Saccharomyces* spp.

Eleanor Harrison

A thesis submitted for the degree of Doctor of Philosophy of Imperial College London

Imperial College London
Silwood Park Campus
Division of Biology

July 2010

Abstract

The 2 Micron plasmid is a multicopy DNA circle inhabiting the genome of the budding yeasts, *Sacchormyces* spp. The plasmid confers no known benefits to the host, but imposes a small fitness cost. However the plasmid is able to drive, i.e. to transmit to >50% of sexual offspring, which allows the element to spread through an outcrossing host population. Therefore we can consider the plasmid a selfish genetic element of yeast.

Here we draw on a number of approaches to improve our understanding of this element. Firstly, we examined the relationship between the cost of plasmid carriage and copy number by experimentally manipulating the number of plasmids in the host. We find that host fitness decreases at a rate of ~0.09% per additional plasmid. Secondly we use experimentally evolving yeast populations to test the hypothesis that sexual reproduction, which is fundamental to the evolution of selfish genetic elements, will drive increasing virulence in the plasmid. We find that 2 Micron copy number increased in outcrossing populations but remained constant in asexual populations. We also find that sex allowed the invasion of non-functional mitochondria in to the populations, showing that sex has the capacity to generate a driving selfish genetic element from one of the most fundamental endosymbionts of the eukaryotic cell.

In addition, we have investigated plasmid variation from global populations of *Saccharromyces spp.* in order to better understand the population biology and evolution of this plasmid. Here we find evidence that the plasmid is able to move between species, recombine with other plasmids within the cell, and exist at a surprisingly wide range of copy numbers in different host populations. Understanding the population structure and evolution of this element allows us to view the plasmid as an autonomous unit evolving in its own right in the genomes of its hosts.

Acknowledgements

A great deal is owed to the patience of my three supervisors Austin Burt, Craig MacLean and Vasso Koufapanou. Thank you for all your help, advice and guidance.

Lots of additional thanks to…

Roshan Vijendravarma for being a fantastic PhD mentor at the start of my degree. I still remember all the good advice you gave me, and I now entirely agree that it would have been a good idea to have taken it.

Matt Goddard for the use of his strains, Paul Sneigowski and Chi-Ren Tien for wild contributing strains, David Clark for providing the pRS-GAL-FLP plasmid, Maitreya Dunham and Matt Hahn for generously contributing their data, Jason Tsai and Aidan Keane for being so very clever and Sue Lomas for all sorts of help!

Silke Fuchs and her lab for letting me impose. I really appreciate your generosity.

Lena, Lynsey and Yael, not that it needs saying. In particular to Lynsey as chief draft reader, Yael for her phylogenetic insights and Lena who, in the twilight of my PhD, as my memory began to go, always knew what I was supposed to be doing when I had no idea.

Mum, Dad and Jane for putting up with me for the worst part of it all.

Thomas Sloan.

Declaration of originality

I declare that all the work presented in this thesis is my own original research with the following acknowledgements for each chapter:

Chapter 2: The pRS-GAL-FLP plasmid was provided by D. Clark, National Institute of Child Health and Human Development. Data on the fitness cost of aneuploidy was provided by M. Dunham, University of Washington.

Chapter 4: Synomymous divergence data for *Saccharomyces* nuclear genes was provided by M. Hahn, Indiana University, and data on codon bias for the same ORFs was taken directly from Hirsh et al. (2005).

Eleanor Harrison

Common abbreviations

2µm	The 2 Micron Plasmid
[cir^0]	2 Micron plasmid free cells
[cir^+]	2 Micron plasmid containing cells
[pRS^0]	pRS-GAL-FLP vector-plasmid free cells
[pRS^0]	pRS-GAL-FLP vector-plasmid containing cells
ORF	Open Reading Frame
LUR	Left Unique Region
RUR	Right Unique Region

Contents

Abstract **2**

Acknowledgments **3**

Declaration of originality **4**

Common abbreviations **5**

1. Introduction **9**

1.1 Sex and conflict 9

1.2 2 Micron Plasmid 10

1.3 Experimental System: Saccharomyces cerevisiae *and* paradoxus 19

1.4 Aims of this thesis 20

1.5 Thesis organisation and content 20

2. The Cost of Copy Number in the 2 Micron Plasmid of *Saccharomyces cerevisiae* **22**

2.1 Introduction 22

2.1.1 Cost of Plasmid Carriage 23

2.1.2 Aims 24

2.1.3 Experimental system 24

2.2 Materials and Methods 24

2.2.1 Strains and culture methods 24

2.1.2 Fitness estimates 26

2.1.3 Copy number analysis 31

2.1.4 Statistical Analysis 31

2.3 Results 32

2.3.1 Copy number change under FLP titration 32

2.3.2 Fitness effects of 2μm copy number changes 33

2.4 Discussion 39

2.4.1 Slow Growth Phenotype 39

2.4.2 Environmental variability in the cost of plasmid carriage 40

2.4.3 Mechanisms for the cost of plasmid copy number 41

2.5 Conclusions 43

2.5.1 Variability in the Cost of Plasmid Carriage 43
2.5.2 Future Directions 44

3. Sex and Selfishness: The Influence of Sexual Strategy on Virulence Evolution in the 2 Micron Plasmid of Saccharomyces cerevisiae **45**
3.1 Introduction 45
3.1.1 Sex and intragenomic conflict 45
3.1.2 Sex and selfish genetic element virulence 46
3.1.3 Aims 47
3.1.4 Experimental system 47
3.2 Materials and methods 49
3.2.1 Strains and culture 49
3.2.2 Analysis of evolved populations: Population level analysis 52
3.2.2 Analysis of evolved populations: Within Population Variation 55
3.2.3 Analysis 55
3.3 Results 56
3.3.1 Copy number evolution 56
3.3.2 The invasion of mitochondrial mutants 61
3.3.3 Fitness 64
3.4 Discussion 70
3.4.1 Sex supports the evolution of selfish genetics elements 70
3.4.2 Sex did not alter the scale of adaptation in evolved populations 73
3.4.3 The extent of selfish genetic element burden does not predict the fitness of evolved populations 74
3.4.4 Antagonistic coevolution negates the benefits of sex 75
3.5 Conclusions 77
3.5.1 Future directions 78

4. Population Biology of the 2 Micron Plasmid **79**
4.1 Introduction 79
4.1.1 2 Micron Plasmid Variation 80
4.1.2 Population Biology of Saccharomyces spp. 80
4.1.3 Aims 81
4.2 Materials and Methods 81

4.2.1 Prevalence 81
4.2.2 Sequencing 82
4.2.3 Copy number 84
4.3 Results and discussion 85
4.3.1 Prevalence and distribution 85
4.3.2 Copy number variation 90
4.3.3 DNA Sequence variation 91
4.5 Conclusions 108
4.A Appendix 111
4.A.1 Complete list of strains used 111
4.A.2 Complete alignment of all plasmid sequences 115

5. Conclusions **127**
5.1 The life and times of the 2µm plasmid 127
5.1.1 Copy number evolution and variation 127
5.1.2 Genome evolution 128
5.1.3 Population biology of the 2µm 129
5.1.4 A final word on the 2µm plasmid 130
5.2 Sex and the evolution of Selfish Genetic Elements 131
5.2.1 Levels of Selection 131
5.2.2 Evolution of Endosymbiosis 131
5.2.3 Evolution of Sex 132
5.3 Conclusions 133

6. Appendix I: General Materials and Methods **134**
6.1 Strains 134
6.2 Microbiology 138
6.3 Molecular biology 139

7. References **144**

Chapter one

Introduction

The amazing complexity of any organism, no matter how basic, is the consequence of a whole suite of genes working intricately together. For the majority of the genome, cooperation between genes is maintained by natural selection, as the inheritance of most alleles is 'fair', i.e. all alleles are transmitted with equal likelihood into the next generation. Therefore, the rise and fall of alleles in the population is due to selection, acting at the level of the individual, to increase the frequency of beneficial, cooperative alleles. However, this system is not foolproof. Where alleles are able to gain a transmission advantage over the rest of the genome, they are able to increase in frequency without relying on selection (Hickey 1982). Thus, alleles that are detrimental to the individual can spread through a population. As these alleles are not contributing to the fitness of the individual, we can refer to them as 'selfish'. Here we define selfish genetic elements as being any genetic unit (from a small non-coding region, to entire chromosomes and organelles) that promotes its own replication without contributing to the fitness of the individual (Burt & Trivers 2006).

1.1 Sex and conflict

Sexual reproduction plays a key role in the evolution of selfish genetic elements. In asexual organisms, the entire genome is inherited as a single unit. Consequently, each allele shares the same evolutionary fate as the whole of the genome, and by extension, the individual. Therefore, the genes of asexually reproducing organisms are dominated by selection acting on the individual. In contrast, under sexual reproduction, the association between alleles is transitory. Through genetic recombination, the linkage disequilibrium within the genome is broken down and alleles are shuffled into novel backgrounds. Thus, alleles, or linked groups of alleles, become the smallest units of selection (Hurst 1992). Therefore, sexual genomes are collectives, in which different genes promote their own agenda. Conflict emerges when the agenda of one part of the genome does not agree with that of the rest (Partridge & Hurst 1998).

However, the vast majority of conflicts between alleles will not be realised, as Mendelian inheritance ensures that the transmission of genes remains 'fair'. Therefore, the frequency of most genes will not increase without positive selection for the phenotype they promote (Crow 1988). Selfish genetic elements are able to overcome this system, and by transmitting to over 50% of meiotic progeny, such genes can spread through the population (Hickey 1982). This over-representation in the offspring of sexual organisms is referred to as drive (Burt & Trivers 2006). In sexual populations, selfish genetic elements which are not only non-beneficial, but incur a fitness cost to the individual, can spread, as long as the rate of drive exceeds the rate of loss through negative selection (Futcher et al. 1988; Hickey 1982). Therefore, an element able to achieve complete drive (d=1) in an entirely outcrossed population has twice the transmission of a regular, Mendelian gene (d=0.5), and can theoretically reduce fitness by up to 50% while still being maintained in the population (Hickey 1982).

Drive

The success of this lifestyle is evident in the huge diversity of elements, employing a wide variety of mechanisms to get around the constraints of Mendelian inheritance (Burt & Trivers 2006). These mechanisms can be summerised into three broad approaches; Firstly, elements can simply over-replicate themselves relative to the genome. This allows elements to drive during segregation. This strategy is employed by numerous elements including the massively successful transposable elements, which cut and paste themselves throughout the genome (Burt & Trivers 2006), homing endonuclease genes, which exploit the DNA repair mechanisms (Burt & Koufopanou 2004), and the 2 Micron plasmid of yeast, which will be discussed in this thesis (Volkert et al. 1989). An alternative approach is to bias transmission after segregation, by inhibiting offspring not containing copies of the selfish element, as is the case in the *t* haplotype in mice or the Segregation Distorter in *Drosophila* (Lyttle 1991). Therefore these selfish genetic elements are present in more than 50% of *viable* offspring. A third strategy is to bias segregation towards the germ line and away from somatic cells, thereby increasing the prevalence of the element in cells contributing to future generations. This strategy is employed by some B-chromosomes, for example by preferentially segregating into the egg rather than the polar body (Jones 1991).

The evolutionary importance of selfish genetic elements

With a few rare exceptions (Gardner et al. 2002), selfish genetic elements can be found in almost every organism studied. Consequently, their role in the evolution of their hosts is

becoming increasingly recognised (Burt & Trivers 2006; Hurst & Werren 2001; Hurst 1992; Johnson 2007). Selfish genetic elements represent a burden to their host populations due to the vast amount of non-beneficial DNA that can accumulate in the host genome. In the human genome, for example, 45% of DNA is estimated to be made up of transposable elements (Hurst & Werren 2001). In addition, the method of drive employed can have major consequences for the evolution of host species. For example, genomic rearrangements caused by transposons may lead to speciation events (Rebollo et al. 2010), and sex ratio distorters that favour males over females have the potential to drive populations to extinction (Lyttle 1977).

Adaptation towards the control of such elements may itself have lead to at least two major advances in eukaryotic evolution. RNAi and methylation, which now play a vital role in gene silencing in eukaryotes, may have evolved originally as a 'molecular immune system' to silence transposons (Johnson 2007), while the evolution of anisogamy (a differentiation between small male and large female gametes) may be an adaptation to prevent conflicts between the cytoplasmic elements of the cell (Hastings 1992; Hurst & Hamilton 1992). Unsurprisingly, there are also many examples of domestication of selfish genetic elements for diverse purposes including roles in the immune system (Agrawal et al. 1998) and mate type switching genes in yeast (Butler et al. 2004).

1.2 2 Micron Plasmid

The 2 Micron plasmid (2μm) of the Saccharomycete (budding) yeasts is a simple, elegant example of a selfish genetic element. Unlike bacterial plasmids, which are often large, functionally significant elements for the host populations (Kado 1998), the 2μm is a small, 6.4Kb, element that encodes just four proteins (Volkert et al. 1989). Each of the plasmid proteins is employed in its own replication, with no beneficial, or actively harmful, functions discovered to date. However, it exists at around 24 – 88 copies per haploid cell (Gerbaud & Guerineau 1980), which amounts to between 1.3 – 4.5% of nuclear DNA, and imposes a growth rate cost, giving yeast cells which lack the element ([cir^0]) a selective advantage of ~1 - 3% (Futcher & Cox 1983; Mead et al. 1986). Despite this cost, the element has been found to be widespread among *Saccharomyces cerevisiae* strains (Futcher et al. 1988), and has also been found in related wild species, *S. paradoxus* and *S. bayanus* (Nakayashiki et al. 2005).

The plasmid's host species' are facultatively sexual, i.e. they can reproduce both sexually and asexually, described in figure 1.1. Under good conditions, diploid cells reproduce

asexually by budding daughter cells from the mother. Under stressful conditions, particularly nitrogen starvation, the cell undergoes sporulation and reproduces meiotically to produce four haploid spores (Bilinski et al. 1986). These spores form a tetrad, which is contained within an ascus cell wall and held together by intra-tetrad bridges (Coluccio & Neiman 2004). The haploid cells of Saccharomyces spp. have two mating types, **a** and α, which are defined by the MAT locus. Matings can occur in three main routes. (a.) The vast majority of mating occurs within the ascus, between members of the same tetrad, and is therefore inbred (Ruderfer et al. 2006; Tsai et al. 2008). (b.) Matings can also take place with spores from other tetrads, which may either be related, and therefore is again inbreeding, or unrelated, and therefore outcrossing. (c.) Haploid spores also have the ability to grow mitotically, during which time some haploid cells are able to switch mating type and mate with their own clone, called autodiploidization (Mortimer et al. 1994). Therefore inter-tetrad mating is the only route by which outcrossed sexual reproduction can occur, however, the clonal nature of yeast means that other spores in the local area are likely to be highly related (Koufopanou et al. 2006). Therefore outcrossing has been estimated as occurring in ~1% of matings in Saccharomyces populations (Ruderfer et al. 2006; Tsai et al. 2008).

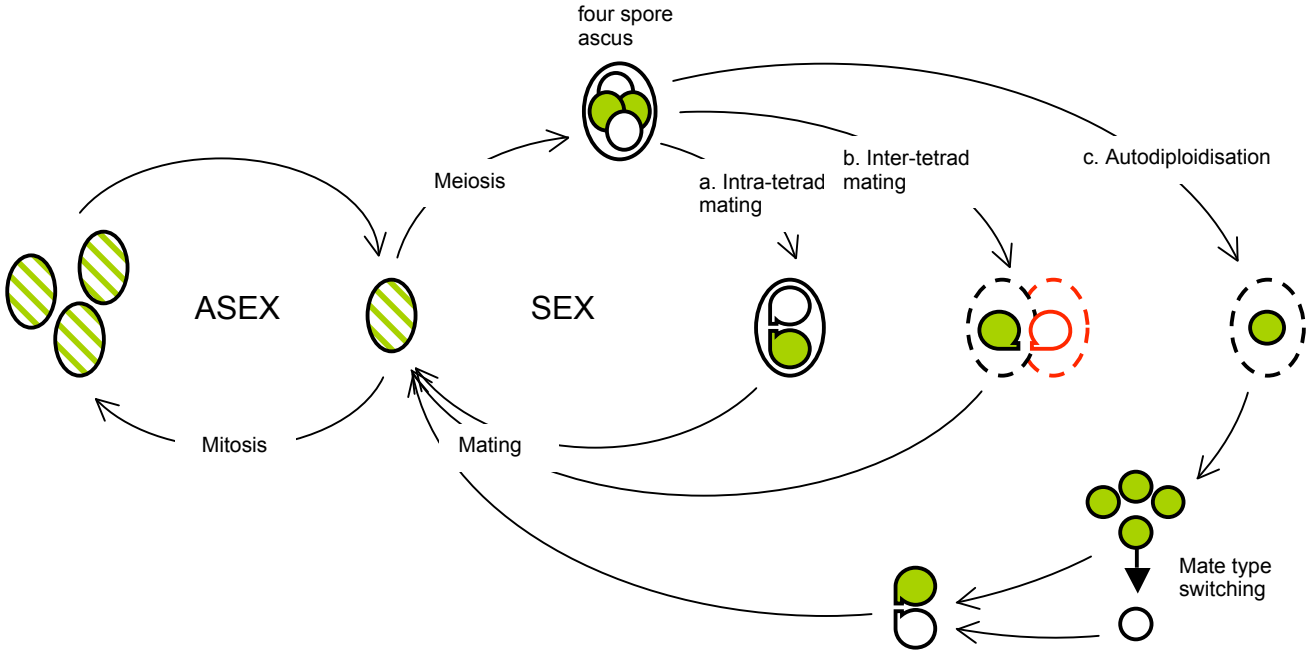


Figure 1.1 *Saccharomyces* spp. lifecycle. Diploid cells (striped) reproduce via mitosis under good conditions. When starved of nitrogen, cells undergo meiosis to produce a four-spore tetrad encased in an ascus cell wall (a tough cell wall produced from the maternal cell wall). Spores produced are of two mating types, a and α, and can either mate with another spore within the ascus (a. intra-tetrad mating), with spores from another tetrad (b. inter-tetrad mating), or reproduce mitotically as haploids before undergoing mate type switching and mating with a haploid clone (c. autodiploidisation).

The relative frequencies of these different reproductive strategies have important consequences for selfish genetic elements of yeast. In the absence of outcrossed sex, the selective advantage of [cir^0] cells leads to a steady loss of plasmid-containing ([cir^+]) cells from the population (Futcher et al. 1988). The plasmid achieves near perfect drive, transmitting to 100% of meiotic progeny (Livingston 1977), with miss-segregation events, which creates [cir^0] cells, occurring at a rate of ~7.6 x 10^{-5} (Futcher et al. 1988). Therefore, despite the selective advantage of [cir^0] cells in the population, the plasmid can be maintained as long as the host population is occasionally sexual. Futcher et al (1988) estimated that just one generation in 5000 is required in order to prevent the loss of this element from the population.

Genome organisation

The tenacity of the plasmid is achieved by just four gene products and four *cis*-acting regions, shown in figure 1.2. The plasmid genome is divided by a pair of identical 599bp inverted repeats, which each contain a 51bp FLP Recognition Target (FRT) (Velmurugan et al. 2003). These regions allow the plasmid to be divided up into a Left and Right Unique Region, each containing two Open Reading Frames (ORFs). In addition, the Right Unique Region also carries two *cis*-acting regions, STB (involved in plasmid segregation) and *Ori* (origin of replication). Between them, these eight regions carry out two functional roles vital to the plasmids maintenance; segregation and copy number amplification.

Plasmid segregation system

The plasmid segregation system shows extremely strong similarities to the cell's chromosome partitioning system, to the extent that they have been suggest as being either related and/or closely coevolving (Ghosh et al. 2010; Malik & Henikoff 2009). The plasmid's STB locus acts much like the centromeric CEN region of the yeast chromosomes. The multiple copies of the plasmid are found clustered into 3 to 5 groups, localised in the nucleus, which then divide just like the sister chromosomes of mother and daugher cells (Jayaram et al. 2004). Segregagation is achieved by the Rep1p and Rep2p proteins, and the STB locus. The Rep1p and Rep2p proteins form a complex, which recruits the centromeric histone, Cse4 to the STB locus (Hajra et al. 2006). This allows the cohesin complex to also be recuited, which pairs the clusters of plasmid molecules and keeps them tethered while they orientate on the spindle fibres (Ghosh et al. 2010; Mehta et al. 2002).

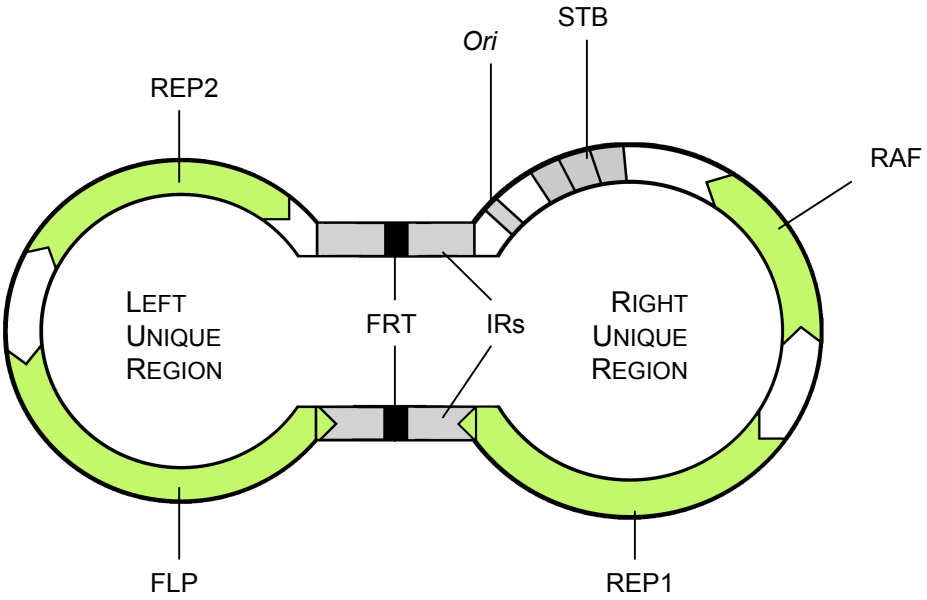


Figure 1.2 Structural organisation of the 2µm genome. The 6.4kb plasmid carries four ORFs (green) and four *cis*-acting regions (grey). The 599bp inverted repeats (IRs) each contain an FRT site (black), which marks the boundary between the two halves of the plasmid, the Left and Right Unique Regions. The FRT sites and the Flpp protein that interacts with them form the basis of the copy number amplification system, while the STB region and Rep1p and Rep2p are used to mimic the cells chromosomal segregation system to ensure the plasmid is segregated between dividing cells (Velmurumgan et al. 2003).

As the cell divides, the cohesin protein is cut, allowing the sister chromatids and plasmid clusters to divide between the cells.

Amplification system

The plasmid segregation system is accompanied by a copy number amplification system, in place to correct for any drop in copy number due to unequal partition of plasmids between dividing cells. Copy number is primarily controlled by the site specific endonuclease, Flpp, which acts at the FRT (Flpp Regognition Target) region. However, all four plasmid proteins ultimately play a role in copy number regulation (Murray et al. 1987).

The plasmid's replication occurs, along with nuclear DNA, only once during the cell cycle (Zakian et al. 1979). Thus, to amplify copy number during this single event the plasmid uses

a method of rolling circle replication (Velmurugan et al. 2003). Flpp recognises the FRT sites contained within the inverted repeat regions. Recombination between these sites allows the plasmid to be replicated multiple times during DNA replication, following the method described in figure 1.3. The expression of this gene is carefully controlled by an antagonistic feedback mechanism, which responds to the concentration of plasmid proteins in the cell (Murray et al. 1987). FLP expression is repressed by Rep1p and Rep2p, therefore the levels of these proteins are sufficient to down regulate FLP and halt the increase in copy number at high copy numbers. The repression of FLP is antagonised by Rafp, which inhibits the Rep1p and Rep2p proteins from interacting with FLP. In addition, the expression of RAF is reduced by the Rep1p-Rep2p complex (Murray et al. 1987)! This seemingly excessive level of repression and inhibition means that copy number recovery can be fast, but remains precise. For instance, when copy number falls, the repression of FLP is lifted in two respects: first, through a reduction in Rep1p and Rep2p inhibition on FLP and second, on RAF, which is then able to further decrease the repression on both genes by Rep1p and Rep2p.

Plasmid copy number is also dependent on the host. Mutations in a number of host genes lead to increasing plasmid copy number, from 2 to 68 times that of the wild type (Burgess et al. 2007; Chen et al. 2005; Dobson et al. 2005; Holm 1982; Sleep et al. 2001; Xiong et al. 2009; Xu et al. 1999). All these genes are associated with the post-translational modification of cellular proteins from two different pathways. First, the E2 ubiquitination pathway, which marks proteins for degradation (Sleep et al. 2001), and second, those from the SUMO (Small Ubiquitin-like Modifier) family which have diverse functions including protein transport, transcription regulation and the cellular stress response (Burgess et al. 2007; Chen et al. 2005; Dobson et al. 2005; Xiong et al. 2009; Xu et al. 1999). This suggests that host mediated destruction of plasmid proteins, most likely Flpp, plays an important role in the control of the 2µm plasmid. The role of the SUMO pathway is, to date, less clear.

Ecology

Approximately 95% of *S. cerevisiae* strains have been found to contain the 2µm plasmid (Xiao & Rank 1993), while it has also been identified in the related wild species *S. paradoxus* and *S. bayanus* (Nakayashiki et al. 2005). However, very little is known about the population structure of the plasmid. Variation in copy number of between 24 and 88 copies per haploid was observed among just four *S. cerevisiae* strains (Gerbaud & Gurineau 1980) and two variants of the plasmid have been identified from industrial *S. cerevisiae*

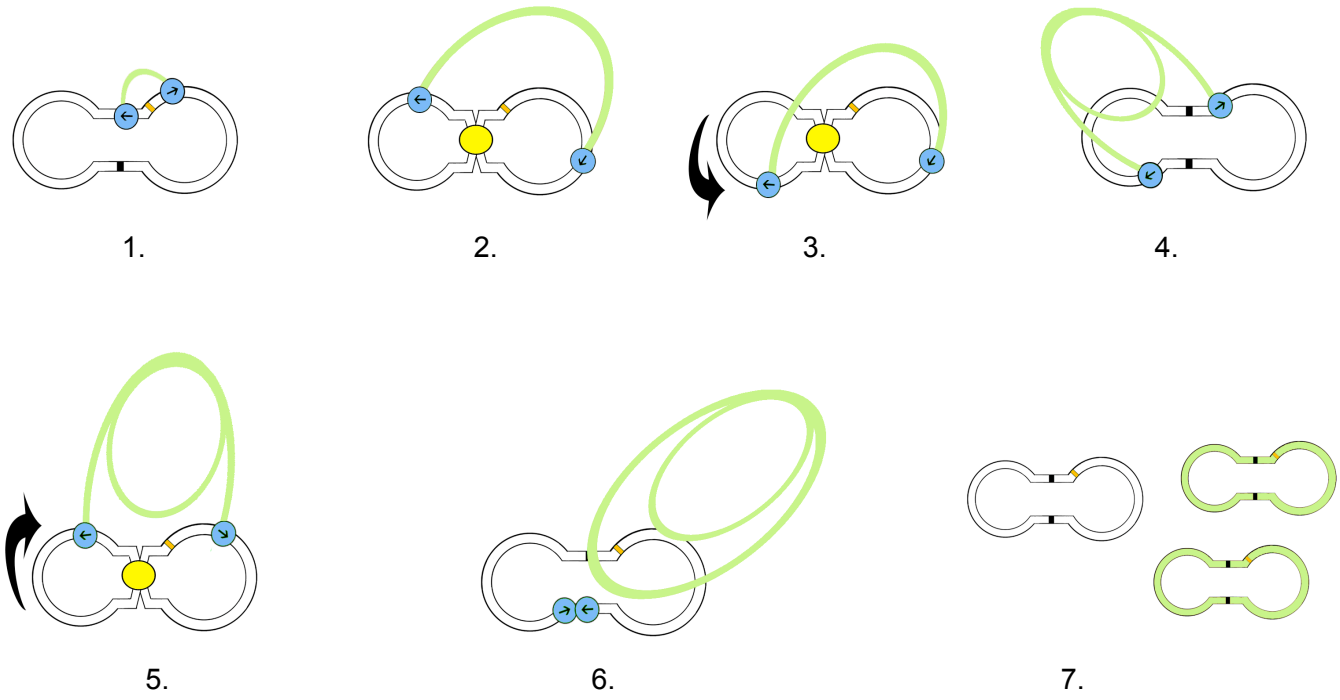


Figure 1.3. Rolling circle amplification of the 2μm plasmid. (1) DNA replication is initiated from the origin of replication (orange) which is positioned next to one of the FRT sites, with replication forks (blue) travelling in opposite directions. The second strand of DNA is shown in green for clarity. (2) Flpp (yellow) causes recombination between the FRT sites, which (3) flips one side of the plasmid relative to the other so that the replication forks are now travelling in opposite directions. (4) The replication forks now chase each other around the plasmid creating a polymer of several plasmid sequences. (5) A second Flpp mediated recombination event flips the plasmid into its original orientation, with replication forks now travelling towards each other once again. (6) Replication is terminated when these forks meet, and the multiplasmid polymer can be resolved into individual plasmids (7). Adapted from Velmurugan et al (2003).

strains based on RFLP and sequence variation (Xiao & Rank 1993). These two variants are characterised by a large degree of divergence (~30%) in the STB region, as well around ~10% amino acid divergence in the RAF and REP1 genes. The 'type 1' plasmids were found almost exclusively in the baking strains of yeast, while 'type 2' plasmids were found distributed among baking, wine, beer, sake and distillation strains. However, with the increasing sensitivity and availability of methods such as quantitative PCR and sequencing, it is likely that there is a great deal more variability yet to be discovered.

Evolution

Despite its success in *S. cerevisiae*, the plasmid is conspicuously absent from other eukaryotic lineages. Just ten nuclear DNA plasmids have been found in eukaryotic species, all of which are found among the Saccharomycetaceae yeasts (figure 1.4), despite a survey of nearly 500 yeast species (Blaisonneau et al. 1997). Each of these plasmids shares the same basic structure as the 2μm; plasmids are small, between 4.7 and 6.7kb, and carry three or four ORFs, divided by two inverted repeat regions, as summarised in table 1.1 (Blaisonneau et al. 1997; Utatsu et al. 1987). Nucleotide homology between these plasmids is negligible, although the proteins do share some local similarities in amino acid sequence (Murray et al. 1988; Volkert et al. 1989). From this, and analysis of function, all the plasmids appear to carry a version of FLP, REP1 and REP2 and consequently use the same segregation and copy number amplification mechanisms (Murray et al. 1988). Therefore, these plasmids clearly represent a related group of elements.

The plasmids appear to have evolved in a host species-specific manner. Plasmid GC content correlates strongly with that of the host species (Blaisonneau et al. 1997) and most are unable to segregate successfully in non-host species, with the exception of plasmids from the most closely related taxa (Murray et al. 1988). In addition, the proteins and *cis*-acting regions of the different plasmids are not interchangeable, indicating co-evolution between the different plasmid loci of the copy number and segregation systems (Murray et al. 1988).

The isolation of these plasmids within the budding yeasts, and their absence from other eukaryotic lineages, appears to be due to their centromeric-like stability system. Members of this yeast family are the only clade known to use short, DNA specified centromeric regions ('point' centromeres), while other eukaryotes use epigenetically defined centromeres, which are not defined by any particular sequence motif (Malik & Henikoff 2009). Therefore, the success of these plasmids within the budding yeast family is due to

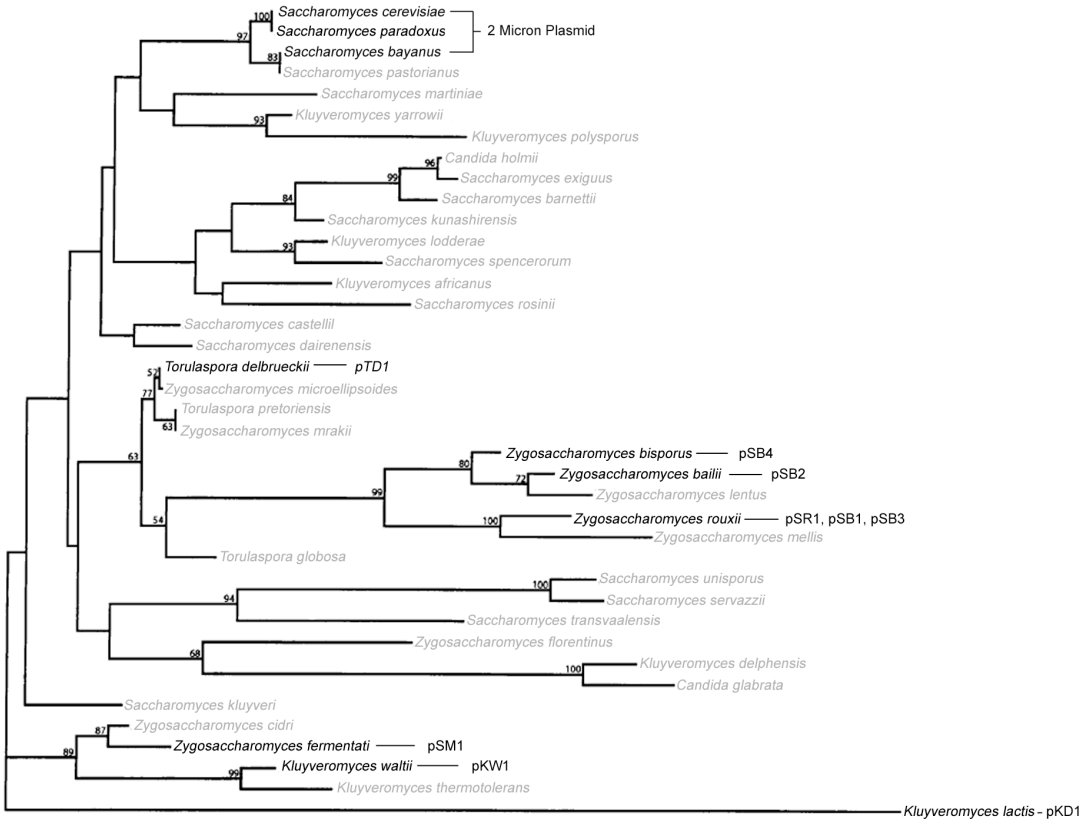


Figure 1.4 Distribution of the 2µm-like plasmids among the Saccharomycetaceae yeasts. Species identified as containing plasmids are shown in black and the name of the plasmid is shown adjacent. Adapted from (Steels et al. 1999).

Species	Plasmid	No. ORFs	(bp)	Reference
Saccharomyces spp	2µm	4	6318	(Hartley & Donelson 1980)
Zygosaccharomyces				
Z. rouxii	pSR1	3	6251	(Araki et al. 1985)
	pSB3	3	6615	(Tohe & Utatsu 1985)
	pSB1	3	~6000	(Tohe et al. 1982)
Z. bailii	pSB2	3	5415	(Utatsu et al. 1987)
Z. bisporus	pSB4	3	~5800	
Z. fermentati	pSM1	4	5416	(Utatsu et al. 1987)
Kluyveromyces				
K. lactis	pKD1	3	4757	(Chen et al. 1986)
K. waltii	pKW1	4	5619	(Chen et al. 1992)
Torulaspora delbrueckii	pTD1	3	4776	(Blaisonneau et al. 1997)

Table 1.1 The 2 Micron-Like plasmids found to date following a survey of over 500 species of yeast. Adapted from Blaisonneau et al. (1997)

their ability to mimic the point centromeres of chromosomal DNA and achieve this extremely stable inheritance. Where centromeres are not sequence-specific, but are instead defined epigenetically by the host, they are vastly more robust to this form of genetic hitch hiking (Malik & Henikoff 2009). Indeed, Malik and Henikoff (2009) argue that the evolution of point centromeres may have occurred on the ancestor of the 2μm. This argument posits that the STB region and the associated ORFs, REP1 and REP2, evolved first to mimic the epigenetic segregation system of their host. These may then have become incorporated into the chromosomal DNA to create the derived point centromeres of the budding yeasts. The presence of 2μm-like plasmid DNA, albeit non-functional, has been observed previously in the genomes of other budding yeasts *Kluyveromyces polysporus* and *Zygosaccharomyces rouxii* (Frank & Wolfe 2009; Utatsu et al. 1988).

1.3 Experimental System: Saccharomyces cerevisiae and paradoxus

The attraction of the 2μm plasmid is, in part, due to its residence in one of the most useful model systems in biology. *S. cerevisiae,* the yeast species used in baking or brewing, benefits from all of the essential traits that make microbial experimental evolution such a powerful approach.

- Short generation times (~1-2hrs) mean that experiments which last only a few weeks can encompass hundreds of generations, allowing evolutionary events to be observed in a relatively short time-scale.

- The small size of yeast cells mean that large population sizes can be maintained, which allows selection to occur in small scale environments. Millions of cells can be maintained in 10ml culture populations, which also allows many replicate populations to be maintained in a relatively small space.

- Yeast can also be stored at -80^{o}C for several years and can be revived for later use. This 'living fossil record' allows the experimenter to directly compare both evolved and ancestral strains under controlled conditions.

In addition, it also comes equipped with powerful molecular protocols for genetic manipulation, complete genome sequences of dozens of strains (Carter 2005), and near complete categorisation of every coding region of its genome (in the form of the *Saccharomyces* Genome Database: http://www.yeastgenome.org/). Furthermore, the wealth of knowledge available on *S. cerevisiae* can also be applied to its sister taxa, *S.*

paradoxus, which shares the vast majority of its genes (Kellis et al. 2003). Unlike *S. cerevisiae*, *S. paradoxus* does not have a history of interaction with human culture, therefore its population structure and evolution is free from this influence. As a consequence, this species is now becoming increasingly important as a model system outside the laboratory (Johnson et al. 2004; Koufopanou et al. 2006; Tsai et al. 2008).

1.4 Aims of this thesis

There is a rich literature dealing with the molecular biology and behaviour of the 2μm plasmid. However, this has not been matched by an equivalent focus on its ecology and evolution as a selfish genetic element in its *Saccharomyces* spp. hosts. This thesis aims to investigate the evolution of the 2μm from two perspectives; the short-term evolution of plasmid virulence, and the long term evolutionary history of the plasmid in its hosts.

1.5 Thesis organisation and content

Chapter 2: The Cost of Copy Number in the 2 Micron Plasmid of Saccharomyces cerevisiae

This chapter addresses how the copy number of the 2μm plasmid alters the fitness of its host. Multiple lines of evidence point to a negative relationship between copy number and the cost of carrying the plasmid, however, to date, this relationship has not been quantified. This relationship has major implications for our view of selection and evolution of plasmid copy number. Here we address this current shortfall by genetic manipulation of plasmid copy number, which allows us to measure the fitness effects of increasing plasmid burden in the host and provides justification for the view of copy number as a 'virulence' trait in the plasmid.

Chapter 3: Sex and Selfishness: The Influence of Sexual Strategy on Virulence Evolution in the 2 Micron Plasmid of Saccharomyces cerevisiae

This chapter examines the fundamental role of sexual reproduction in driving intragenomic conflict in the cell. As we have discussed, sex (and in particular, outcrossing) provides the basis for the evolution of selfish genetic elements by breaking down the linkage between genes, and through meiosis, provides the means for some genetic elements to gain a transmission advantage over the majority of the genome. Sex also reduces the relatedness among genes, and therefore increases the competition between rival elements within the

genome. We use an experimental evolution approach to test the hypothesis that increasing levels of sex will drive increasing levels of virulence (i.e. copy number) in the plasmid. We also show that sexual reproduction drives the evolution of selfishness among mitochondria. This chapter tests both a fundamental prediction of selfish genetic element evolution, and also provides insights into the forces shaping the evolution of the plasmid itself.

Chapter 4: Population Biology of the 2 Micron Plasmid

Here the evolution and ecology of the plasmid is addressed in its host populations. Very little is known about the distribution and variation of 2μm in global yeast populations, and in particular the distribution of the element in the wild yeast *S. paradoxus.* We use data obtained from both industrial and wild populations of *S. cerevisiae* and *S. paradoxus* to ask questions about its natural ecology and evolution. The aim is to build a picture of the plasmid populations; either as benign passengers in their host lineages or transient elements undergoing continuous loss and invasion cycles in host populations.

Chapter two

The Cost of Copy Number in the 2 Micron Plasmid of *Saccharomyces cerevisiae*

2.1 Introduction

The 2 Micron plasmid (2μm) is a selfish genetic element found widely amongst strains of *Saccharomyces cerevisiae*, and in related taxa, *S. paradoxus* and *S. bayanus* (Nakayashiki et al. 2005). No functional benefit of carrying the plasmid has yet been found, and the detailed descriptions of gene function (reviewed in Chapter 1) leave little likelihood for beneficial functions to be identified in the future. Instead, there is a small but significant cost to carrying the plasmid, measured as between 1 and 3% growth rate reduction in three different haploid strains (Futcher & Cox 1983; Mead et al. 1986). The plasmid is able to persist because of its ability to drive; i.e. to be transmitted to >50% of meiotic progeny (Burt & Trivers 2006). This allows it to spread though sexually reproducing populations, countering the spread of plasmid free ([cir^0]) strains into the population (Futcher et al. 1988).

The copy number of the 2μm is often cited as being ≈60 molecules per haploid genome, however this is highly variable between strains. Previous surveys of *S. cerevisiae* strains have found copy numbers of between 24 and 88 per haploid (Futcher & Cox 1984; Gerbaud & Guerineau 1980), and we have found copy numbers varying from 2 to 753 per diploid genome, in a comprehensive survey of wild, clinical and industrial strains of *S. cerevisiae* and *S. paradoxus* (Chapter 4). As the plasmid is 6.4kb in size, this variation corresponds to between ~13kb and ~4,800kb of extra DNA being supported by the cell, making up over a quarter of the total nuclear DNA in the most extreme case.

Plasmid copy number is regulated by both the host genome and the plasmid itself. DNA replication is initiated only once during the cell cycle along with the rest of the genome, however the plasmid employs a method of rolling circle replication (*see* Chapter 1), mediated by the plasmid encoded Flpp protein, which allows multiple copies to be

produced during a single replication event. However, over replication of the plasmid is prevented by a negative feedback loop involving the remaining three plasmid proteins, which act to control the expression of FLP (*see* Chapter 1 for more details; Murray et al. 1987).

The role of the host in controlling 2µm copy number is demonstrated by a number of *S. cerevisiae* mutant strains, which have elevated plasmid copy number (Burgess et al. 2007; Chen et al. 2005; Dobson et al. 2005; Holm 1982; Sleep et al. 2001; Xu et al. 1999). The genes in question are all involved in the post translational modification of proteins, including genes from the E2 ubiquitination pathway, which encode tagging proteins that mark unwanted or damaged cellular proteins for degradation (Sleep et al. 2001), as well as those from the SUMO (Small Ubiquitin-like Modifier) family which have diverse functions including protein transport, transcription regulation and the cellular stress response (Burgess et al. 2007; Chen et al. 2005; Dobson et al. 2005; Sleep et al. 2001). Elevated copy numbers range from as low as two times (Dobson et al. 2005), up to 68 times (Xu et al. 1999) that of wild type strains, dependant on the mutation. The role that these proteins play in the maintanance of copy number is not clearly understood, although one simple regulatory mechanism is the use of the E2 ubiquitin pathway to mark plasmid encoded proteins for rapid degredation (Sleep et al. 2001).

2.1.1 Cost of Plasmid Carriage

The most simple relationship to explain the cost of plasmid carraige would be that the fitness cost of carrying the plasmid is positively correlated with copy number. This has been observed in the majority of plasmid-bacterial associations, in which there is a clear positive association between copy number and fitness cost (Patnaik 2000). The tight regulation of 2µm copy number by both host and plasmid suggests that this is also the case in this system. This hypothesis is supported by the severe reductions in fitness observed in the high copy mutant strains, which experience not only a reduction in growth rate, but spontaneous cell death, irregularities in cell size and cold sensitivity (Chen et al. 2005; Dobson et al. 2005; Xu et al. 1999). This pathology is not observed when strains are cured of the plasmid, indicating that it is the high copy number of the plasmid, and not the chromosomal mutation, which causes this reduction in fitness. However, the possibility that the fitness costs are associated with an epistatic interaction between chromosomal and plasmid genes, rather than the copy number itself, cannot be excluded.

2.1.2 Aims

This study aims to test the prediction that the fitness cost of plasmid carriage relates to the copy number of the element. Here we investigate the fitness cost of increasing plasmid copy number by taking advantage of the plasmid's own copy number regulation system to manipulate plasmid burden.

2.1.3 Experimental system

We have used a vector-based system that exploits the plasmids own copy number amplification to allow us to experimentally manipulate copy number. The FLP gene from the 2µm is placed under the control of a galactose-inducible promoter on a single copy plasmid vector, pRS-GAL-FLP (Alfieri & Clark 1999). Growth on galactose triggers the over expression of Flpp in the cell, causing the 2µm to reach increasingly high copy numbers. Such constructs have been used regularly to cure strains of the 2µm, on the principle that high copy number is so costly, that any spontaneous plasmid free ([cir^0]) strains have an exaggerated selective advantage and are enriched in the population (Alfieri & Clark 1999; Rose & Broach 1990; Storici & Bruschi 1997; Tsalik & Gartenberg 1998).

The pRS plasmid was used to cure strains of the 2µm, creating a set of strains both with and without these two plasmids. We were then able to control the expression of the FLP gene from the pRS plasmid by growing strains on different concentrations of galactose media. This system has allowed us to amplify copy number up to 10 times the wild type state in this strain, and measure the impact this has on fitness.

2.2 Materials and Methods

2.2.1 Strains and culture methods

Strain construction

Four strains based on diploid strain YDG598 (MAT**a**/MATα *ho/ho his/his leu/leu ura/ura* [cir^+]) were used;
(a) YDG598 [cir^+] [pRS^0], (b) YDG598 [cir^+] [pRS^+], (c) YDG598 [cir^0] [pRS^+] and (d) YDG598 [cir^0] [pRS^0]. The pRS-GAL-FLP plasmid was kindly donated by J. Clark. YDG598 [cir^+] [pRS^0] represents the wild type state of this strain, carrying the 2µm

plasmid (cir) but lacking the vector-plasmid (pRS). The remaining strains were created in following the steps described in figure 2.2.1.

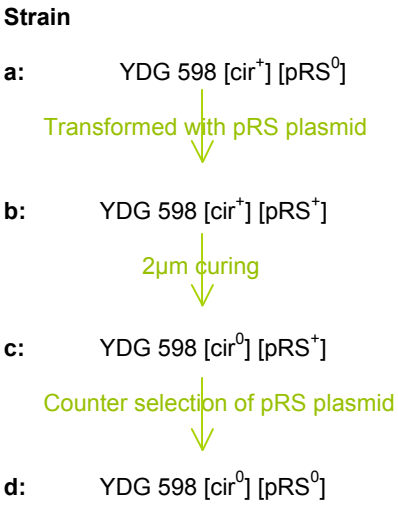


Figure 2.2.1. Steps involved in making four strains used in this study.

Transformation

Four colonies of YDG598 [cir^{+}] [pRS^{0}] were picked and transformed with pRS-GAL-FLP using the high efficiency transformation technique described in Amberg (2005), pp. 109-111. The plasmid carries the URA3 marker, therefore transformants were selected on SD agar lacking uracil (reagents are described for all medias and solutions in appendix I).

Curing

Strains lacking the 2µm plasmid ([cir^{0}]) were obtained by prolonged overexpression of FLP from the pRS vector-plasmid (Alfieri & Clark 1999). Ten YDG598 [cir^{+}] [pRS^{+}] colonies were grown in 10ml SD-GAL media lacking uracil overnight at 30^{o}C. 50µl were used to inoculate fresh media and again grown overnight. This was repeated for a total of four transfers to allow enrichment of [cir^{0}] cells. Cultures were then plated onto solid media (SD-GAL-agar – uracil) and 5 large colonies picked from each plate for 2µm screening.

To identify [cir^{0}] isolates, colony PCR was performed initially as described in appendix I. Colonies showing no amplification of the plasmid target were then picked and grown overnight in 1ml SD – uracil and the DNA extracted for standard PCR analysis using two primers pairs; REP1 (for = TCACCCCACAATCCTTCATC and rev = CATCTGGCCCAAACTTCTC) and REP2_2 (for = GCTGGTGGGACTAATAACTGTG and rev = ACGACATTGAAACAGCCAAG), and genomic control primer pair ITS (1 = TCCGTAGGTGAACCTGCGG and 4 = TCCTCCGCTTATTGATATGC). Colonies negative for both 2µm targets and positive for the genomic control were considered [cir^{0}]. The resulting strain is therefore YDG598 [cir^{0}] [pRS^{+}].

Counter selection of the pRS-GAL-FLP plasmid

The final stain, YDG598 [cir^{0}] [pRS^{0}], which lacks both 2µm and the vector-plasmid, was created by inoculating YDG598 [cir^{0}] [pRS^{+}] colonies into 10ml non selective media (MM + uracil, histidine and leucine) and allowed to grow overnight. The cultures were then

plated onto MM + uracil, histidine and leucine + 5FOA agar to select for loss of the pRS plasmid (as 5FOA is counter selective for the URA3 gene), and streaked onto SD agar – uracil to ensure that they are unable to grow in the absence of uracil.

2.2.2 Fitness estimates

Growth rate analysis was conducted in a Biotek Synergy HT plate reader in 96 well plates. Strains were grown in 200µl of SC media, non-selective media for [pRS^0] strains (a and d; figure 2.2.1) or SD –ura, URA selective media for [pRS^+] strains (b and c; figure 2.2.1) containing 0, 2, 5, 10 and 20g/L galactose as the sole carbon source to control expression of the FLP gene. Five replicate plates were run containing four replicates of each strain in each medium, blanks, and SC + glucose media positive growth controls for each strain. On each plate [cir^0] and [cir^+] strains were paired and placed in adjacent wells. The plates were split into four replicate blocks and the positions of the pairs within each block was randomised. Growth rate was measured by reading turbidity at OD 600nm relative to a reference well (containing media only) every 2hrs for a total of 48hrs at 30^{o}C. All growth curves are shown in figure 2.2.2.

Fitness was calculated using the maximum growth rate for each culture. The plate reader software, Gen5, calculates the maximum linear slope, in mOD minute^{-1}, for each growth curve by sequentially calculating the slope across five adjacent 2 hour time points and returning the maximum value (maxV). As turbidity is proportional to cell density, maxV can be used as an estimate of growth rate (µ) in the equation $N_t = N_0 e^{\mu t}$ (Hartl & Clark 1997), where N_0 is the population size at generation 0, N_t is the population size at time t and e is a constant.

To represent the fitness cost of carrying the 2µm, relative fitness (w) was calculated for each [cir^+] culture compared to its adjacent [cir^0] culture using the equation (Hartl & Clark 1997):

$$w = e \left(\frac{\mu_{cir+}}{\mu_{cir0}} - 1 \right)$$

In addition to fitness measurements, the density at stationary phase was measured as the maximum OD 600nm (maxOD) recorded in each growth curve.

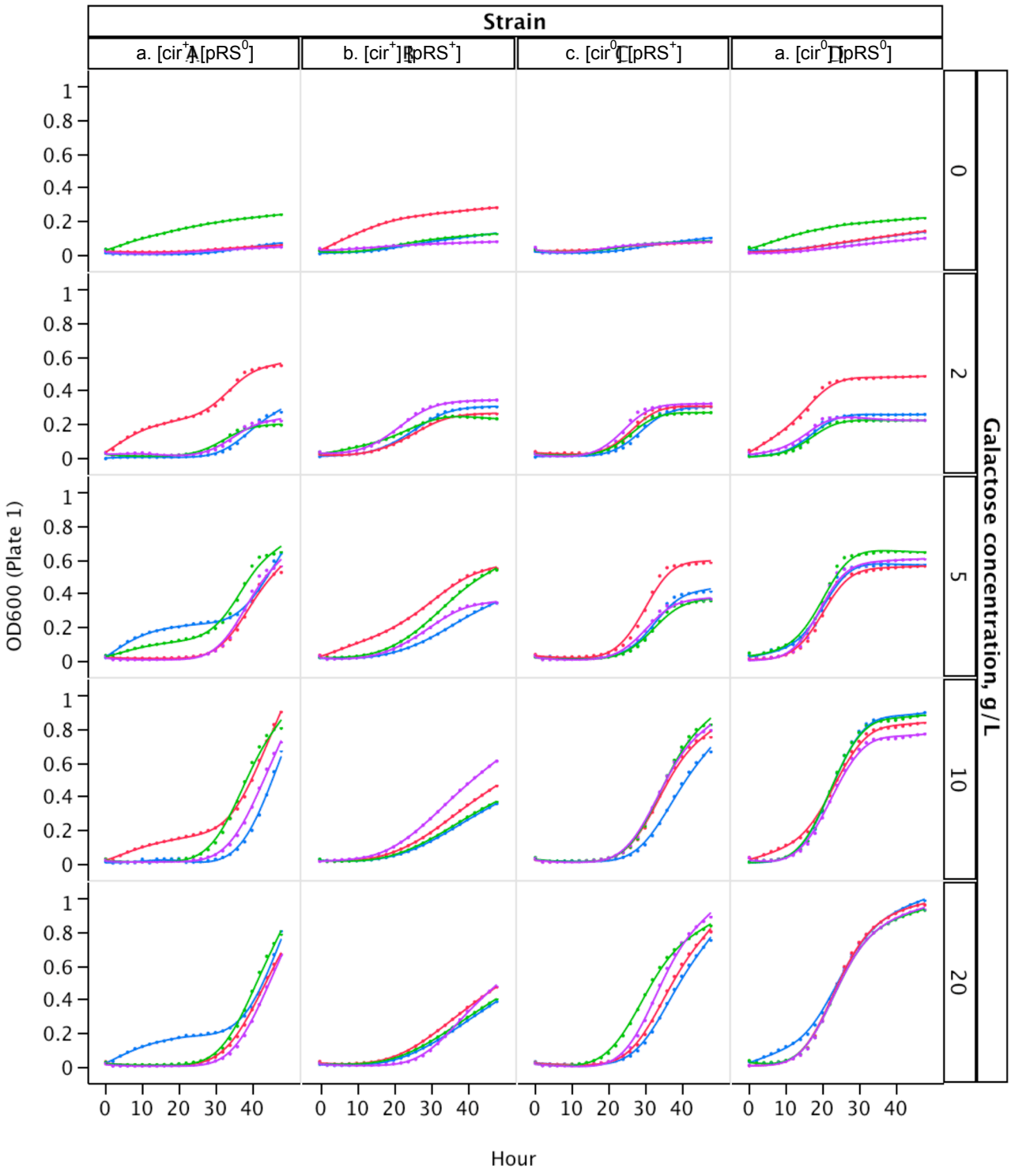


Figure 2.2.2a Growth curves from replica Plate 1. Absorption at 600nm (OD600) was measured in 200µl cultures every 2 hours for 48hrs. Growth plates contained four replicates of each treatment organized into blocks 1 (blue), 2 (red), 3 (green) and 4 (purple). Smoother curves are fitted at lambda 100.

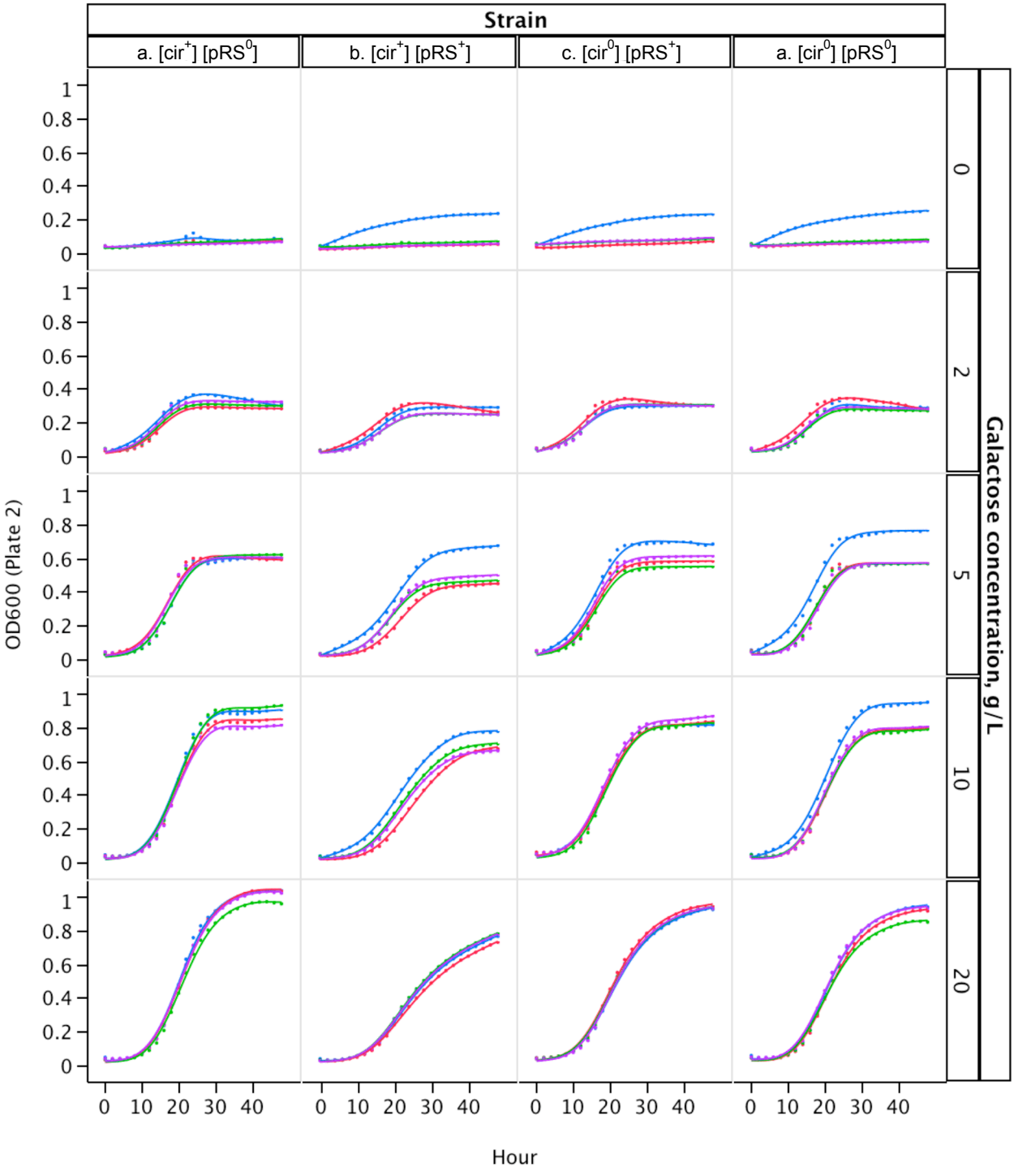


Figure 2.2.2b Growth curves from replica Plate 2. Absorption at 600nm (OD600) was measured in 200μl cultures every 2 hours for 48hrs. Growth plates contained four replicates of each treatment organized into blocks 1 (blue), 2 (red), 3 (green) and 4 (purple). Smoother curves are fitted at lambda 100.

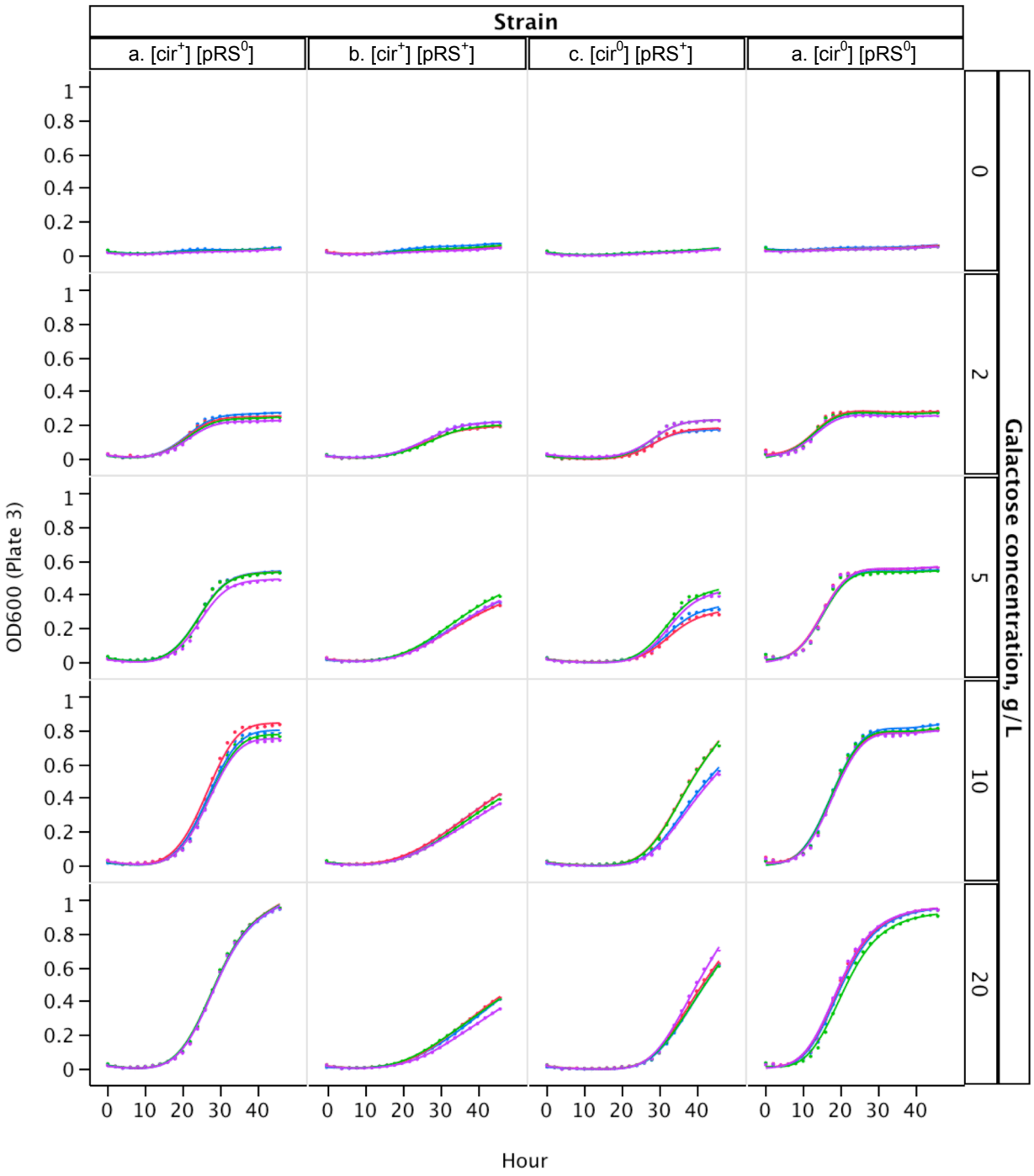


Figure 2.2.2c Growth curves from replica Plate 3. Absorption at 600nm (OD600) was measured in 200μl cultures every 2 hours for 48hrs. Growth plates contained four replicates of each treatment organized into blocks 1 (blue), 2 (red), 3 (green) and 4 (purple). Smoother curves are fitted at lambda 100.

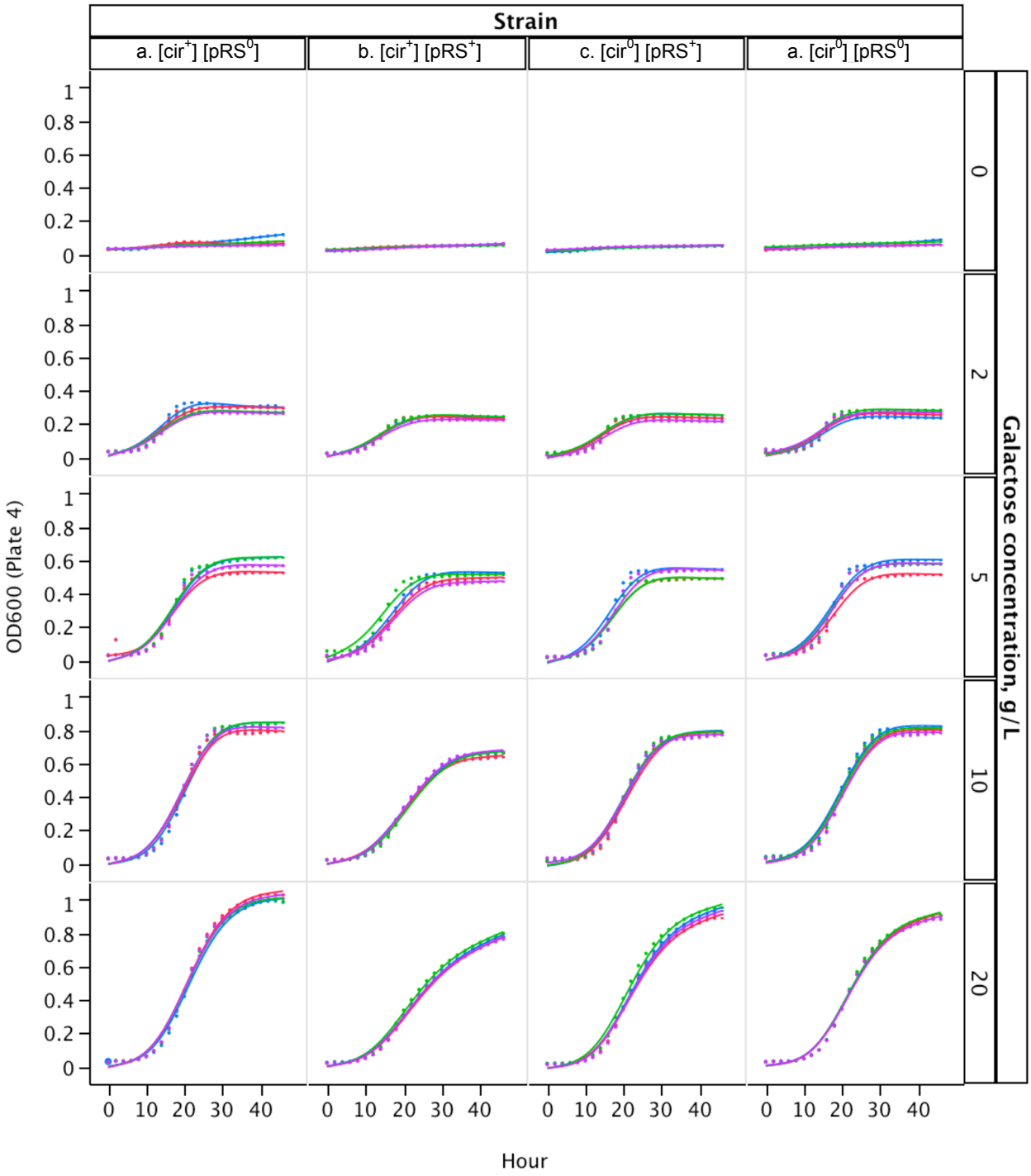


Figure 2.2.2d Growth curves from replica Plate 4. Absorption at 600nm (OD600) was measured in 200µl cultures every 2 hours for 48hrs. Growth plates contained four replicates of each treatment organized into blocks 1 (blue), 2 (red), 3 (green) and 4 (purple). Smoother curves are fitted at lambda 100.

2.1.3 Copy number analysis

Growth plate 5 was removed from the plate reader after 30 hrs, whilst cultures were in mid-log phase. DNA was extracted using the Promega Wizard® SV 96 genomic DNA purification system. Prior to extraction, media was removed by pipetting and cultures were resuspended in lyticase buffer (appendix I). These were then incubated over night at 37^0C and followed by extraction of the DNA following the kit protocol. Quantitative PCR was performed on [cir^+] strains, using the methods described in appendix I. DNA extracts were diluted 10 fold prior to qPCR and DNA from stocks of wild-type strain YDG598 [cir^+] [pRS^0] was used to create standard curves of six 10 fold dilutions from 100ng per reaction. Genomic and 2µm targets were detected by primers qHO (for = CGGTTTGTGACGATCACGTT and rev = CCTGTTTTCATCTGGACCATCTC) and qSTB (for = CATCCCCGGTTCATTTTCTG and rev = AGATATGCTATTGAAGTGCAAGATGG) respectively. Estimated copy number was calculated as described in appendix I.

2.1.4 Statistical Analysis

Where appropriate, data was analyzed using Analysis of Variance (ANOVA) with factors as described in the text, including experimental design effects of plate and block where appropriate, and all interactions between them. In all cases, the continuous variables (galactose or copy number) were treated initially as categorical variables, and then the degrees of freedom divided between the linear function, and the deviation from it (Sokal & Rohlf 1981). Where a linear fit was not appropriate the factor was treated as continuous and a curved function tested. Models were simplified by removing non-significant factors, until the minimum model was reached, using a threshold of $p<0.05$. All analysis was performed in JMP 8 (SAS Institute Inc, Cary, NC, USA).

In order to investigate the effect of copy number on fitness, copy number estimates from plate 5 were applied to the rest of the growth plates. Estimates for each galactose concentration were given as the average for that group. Due to the need to harvest cells prior to saturation phase for DNA extraction, plate 5 was excluded from the analysis of fitness as cultures may not have reached maximum growth rate.

2.3 Results

2.3.1 Copy number change under FLP titration

In order to manipulate the copy number of the 2µm plasmid we transformed strains with a plasmid vector carrying the 2µm FLP gene under the control of a galactose inducible promoter. Copy number could then be controlled by growing strains on different concentrations of galactose media. We tested the change in copy number associated with galactose concentration using quantitative PCR.

Both log10[galactose] and pRS treatment had significant effects on copy number, with a strongly significant interaction between the two (table 2.3.1; ANOVA pRSxlog10[galactose]: $F_{3,24}$=8.011, p=0.0007, N=32). Figure 2.3.1 shows that there is a difference in response to galactose concentration between pRS treatments. Analysis of each strain was then conducted independently (see table 2.3.2). In the [pRS^0] strain there was no effect of log10[galactose concentration] on copy number (ANOVA log10[galactose]: $F_{3,13}$=2.572, p=0.119, N=16). In the [pRS^+] strain there was a strong linear relationship between the plasmid copy number and log10[galactose] (ANOVA log10[galactose]: $F_{1,13}$=124.75, p>.0001, N=16).

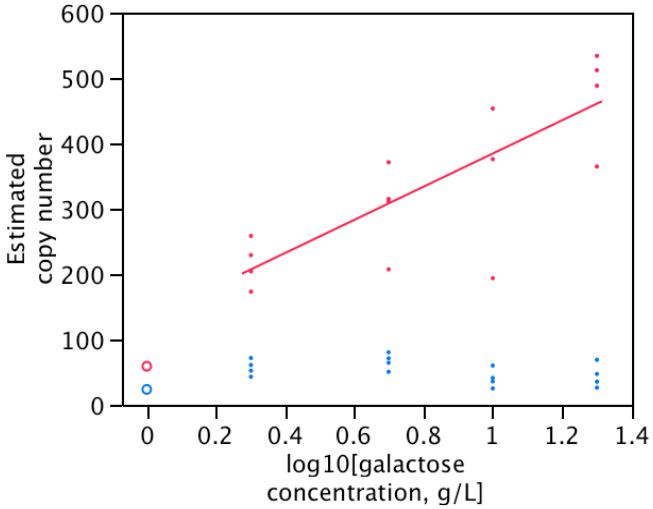


Figure 2.3.1. Effect of galactose concentration on 2µm plasmid copy number in [cir^+] strains either with ([pRS^+]; red) and without ([pRS^0]; blue) the pRS vector-plasmid. Each point represents one replicate culture. Empty circles denote control cultures grown on 20g/L glucose with no galactose, and were not included in the analysis. A significant linear relationship was found between copy number and log10[galactose] in the pRS^+ treatment ($F_{1,12}$=124.750, p>.0001, N=16).

Source	DF	MS	*F*	*p*
pRS	1	663763.8	193.041	<.0001*
log10[galactose]	3	20773.6	6.041	0.0032*
pRS*log10[galactose]	3	27545.5	8.011	0.0007*
Error	24	2996		
			N=32, *F*=33.5998, *p*<.0001	

Table 2.3.1 ANOVA of all significant main effects and interaction effects on copy number.

		A [pRS^0]			**B [pRS$^+$]**		
Source	**DF**	**MS**	***F***	***p***	**MS**	***F***	***p***
log10[galactose]	3	569.510	2.652	0.0962	47749.63	7.167	0.0052*
Linear regression	1	741.032	1.532	0.240	140988.58	124.75	>.0001*
Deviation from linear	2	483.748	2.184	0.155	1130.15	0.170	0.845
Error	12	214.720	6662.2				
		N=16, *F*=2.652, *p*=0.0962			**N=16, *F*=7.1672, *p*=0.0052**		

Table 2.3.2 ANOVA table of the effect of log10[galactose concentration] and replica block on copy number within the two strain treatments, (a.) pRS^0 and (b.) pRS$^+$. Galactose concentration was initially treated as a categorical variable, and then this variation divided up into to that fitting a linear regression and the deviance from that line, following the methods of Sokal and Rohlf (1981).

2.3.2 Fitness effects of 2µm copy number changes

The fitness cost of carrying the 2µm plasmid was calculated by comparing growth rates of [cir$^+$] strains to their [cir^0] partner, placed adjacently in the 96-well plate, calculated as described (Materials and Methods). The use of growth rate to define fitness was validated by comparing maxV to the estimate of turbidity at saturation, maxOD. The two estimates were compared by taking the residuals from an ANOVA with galactose, pRS and 2µm plasmid treatment and plate as main effects, to avoid confounding effects of these factors. Correlation between the two was highly significant (Pearson product moment correlation: r=0.768, p<.0001, N=256; figure 2.3.2), with an estimated slope of 1.486 (95% CIs = 1.342 – 1.65; table 2.3.3).

Owing to our use of galactose concentration to manipulate copy number, these two factors are necessarily confounded, i.e. in the [cir$^+$] [pRS$^+$] strain, copy number and galactose concentration co-vary. In the control, [cir$^+$] [pRS^0] strain, copy number was not effected by galactose concentration, therefore we have used the results of the [cir$^+$] [pRS$^+$] strain to estimate the effect of copy number on fitness, but also analysed the confounding effects of galactose in the [cir$^+$] [pRS^0] strain.

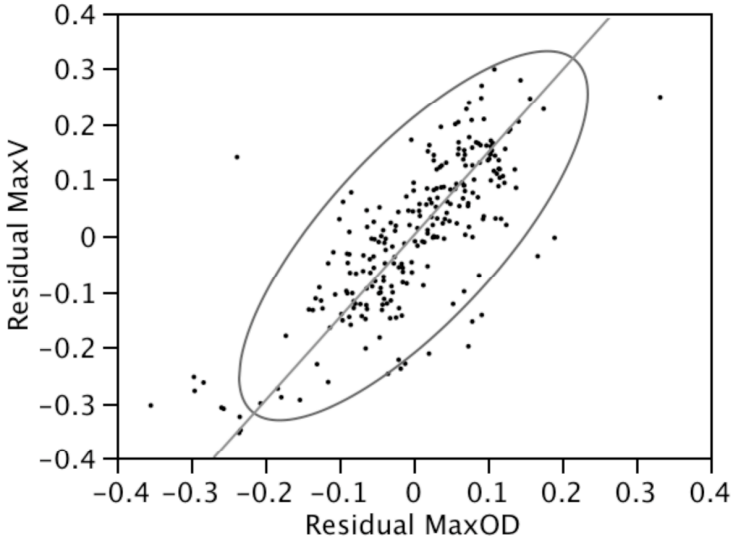


Figure 2.3.2. Pearson product moment correlations between measures of growth rate (maxV) and measures of saturation density (maxOD). Confounding effects of galactose, pRS and 2µm treatments and growth plate were removed by taking the residuals from an ANOVA with these as main effects. 95% density ellipse is shown and the line of fit plotted using Reduced major axis regression, with the variance ration defined as the standard deviation of X over standard deviation of Y.

X	Y	Variance Ratio	Intercept	Slope	Lower 95%	Upper 95%	r	*p*
MaxOD	MaxV	1.427	4.28e-17	1.486	1.342	1.65	0.768	<.0001*

N=256

Table 2.3.3 Comparison of MaxV (growth rate) and MaxOD (saturation density) for their use as fitness estimates. Residuals from an ANOVA with pRS, galactose and plate were compared to remove the confounding effects of these factors. Significance was estimated using Peason's product-moment correlation (r), and the slope fitted using Reduced major axis regression. Variance ration was calculated as the ratio between the standard deviations of the X and Y variables.

Copy number

Estimates of copy number in YDG598 [cir$^+$] [pRS$^+$] at each galactose concentration were averaged and applied to the relevant group of fitness estimates. As the specific cultures used to calculate copy number were harvested in mid-growth phase and not grown to saturation, these could not be included in fitness analysis. For analysis, copy number was originally treated as a categorical variable, and the degrees of freedom from this (df=3) also divided into the linear component (df=1) and deviation from linear (df=2; Sokal & Rohlf 1981). Copy number had a significant main effect on fitness, both as a linear effect (ANOVA copy number(linear): $F_{1,57}$=11.961, p=0.001, N=64) and the deviation from it (ANOVA copy number(deviation): $F_{2,57}$=7.487, p=0.0013, N=64; table 2.3.4a).

To calculate the fitness cost per plasmid, we regressed relative fitness against copy number, constraining the intercept of the regression to 1 (as, according to the assumptions of our fitness equation, a [cir$^+$] strain at a copy number of 0 = a [cir^0] strain). The slope of this regression is –0.0009 (s.e.=3.454x10^{-5}, N=102; figure 2.3.3; table 2.3.4b), implying that each additional plasmid reduces fitness by 0.09% (±0.0003).

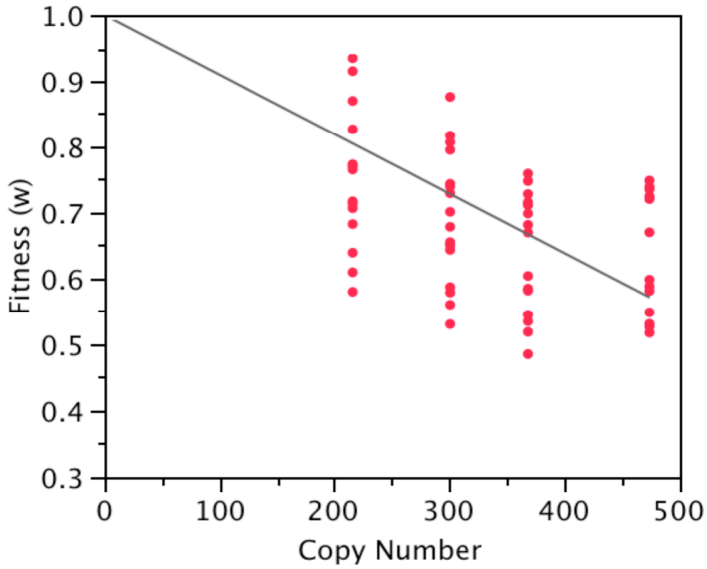


Figure 2.3.3 Relationship between 2µm copy number and fitness in the [pRS$^+$] strains. Copy number estimates were taken from the average of four replicate cultures grown under identical conditions. Each point shows fitness estimates calculated from a single [cir$^+$] [pRS$^+$] strain compared to its [cir^0] [pRS$^+$] counterpart. The regression line was constrained to an intercept of 1 and has a slope of –0.000903 (s.e.=3.454e-5, N=102).

a.	Source	DF	MS	*F*	*p*
	Plate	3	0.0368	6.450	0.0008*
	Copy number	3	0.199	34.843	<.0001*
	Linear regression	1	0.511	11.961	0.0010*
	Deviation	2	0.0427	7.487	0.0013*
	Error	57	0.00571		
	N=64, *F*=20.646, p=<.0001				

b.	Term	Estimate	Std Error	t Ratio	*p*
	Intercept = 1	1			
	Copy Number	-0.000903	3.454e-5	-26.13	<.0001*
	N=102				

Table 2.3.4 ANOVA table for the effect of copy number on fitness in pRS^+ strains. Copy number was initially treated as a categorical variable (Sokal & Rohlf, 1981) (a). To estimate the cost of the plasmid per additional molecule a linear regression between fitness and copy number was fitted, with intercept constrained to 1 (b).

Galactose

To test for the effects of galactose independently of copy number variation we compared the fitness of strains not containing the pRS vector-plasmid. The [pRS^0] control treatment showed a significant positive linear relationship between fitness and galactose concentration (ANOVA galactose(linear): $F_{1,57}$=5.304, p=0.0249, N=64), but also showed a large amount of variation from this linear trend (ANOVA galactose(deviation): $F_{2,57}$=5.706, p=0.0055, N=64; figure 2.3.4; table 2.3.5a). Mean fitness estimates varied between 0.971 (s.e.=0.0415, N=16) at 2g/L galactose up to 1.122 (s.e.= 0.0402, N=16) at 20g/L. This is surprising as it indicates a ~12% fitness advantage of having the 2µm at high galactose concentration. However, this has been previously reported in strains transformed with another 2µm curing plasmid, and appears to be due to a mutation caused during the transformation process (Mead et al. 1987). Therefore, the apparent increase in fitness is most probably due, not to the 2µm, but the effect of a deleterious mutation in the cured strain.

Over expression of the FLP protein

A possible mechanism behind the cost of plasmid carriage is the increasing cost of gene expression. To test this hypothesis we estimated the relative fitness of the [cir^0] [pRS^+] strain, compared to YDG598 [cir^0] [pRS^0], therefore testing the effect of the pRS plasmid in the absence of the 2µm. As pRS-GAL-FLP is a single copy plasmid, changing

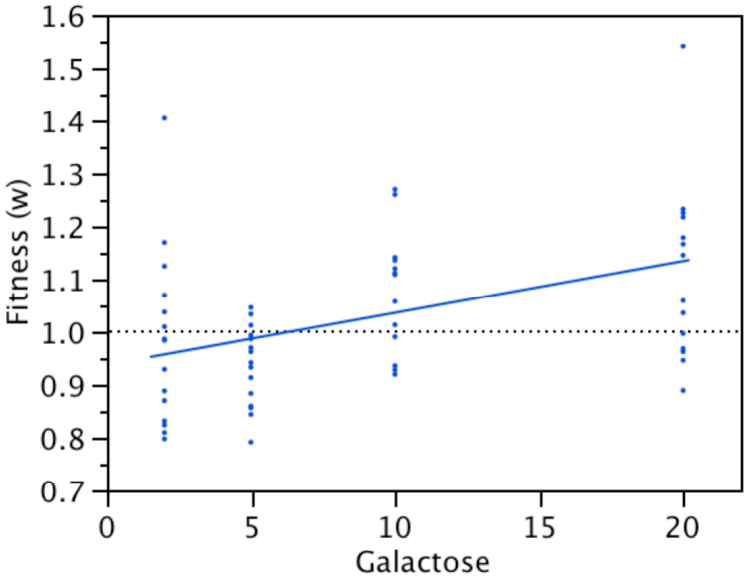


Figure 2.3.4 Effect of galactose concentration on the fitness of 2µm containing strains relative to 2µm free strains, in the absence of the pRS vector-plasmid. Each point is a single estimate of fitness based on one [cir^{+}] [pRS^{0}] and one [cir^{0}] [pRS^{0}] culture grown adjacent to each other in a 96 well plate. Solid lines shown the best fit as determined from the analysis and dotted lines show 95% CI.

a.	Source	DF	MS	*F*	*p*
	Plate	3	0.0842	13.893	<.0001*
	Galactose	3	0.118	19.488	<.0001*
	Linear regression	1	0.183	5.304	0.0249*
	Deviation	2	0.0346	5.706	0.0055*
	Error	57	0.00606		
	N=64, *F*=16.691, p=<.0001				

b.	Source	DF	MS	*F*	*p*
	Plate	3	0.159	14.416	<.0001*
	Galactose	1	0.216	19.620	<.0001*
	Galactose^2	1	0.00776	0.705	0.4046
	Error	58	0.0110		
	N=64, *F*=14.048, p=<.0001				

Table 2.3.5 ANOVA table for the effect of galactose on fitness within the [pRS^{0}] treatment. Galactose was initially treated as a categorical variable and the variance then divided between a linear fit and the deviance from it (Sokal & Rohlf, 1981) (a), Analysis was then conducted using galactose as a continuous variable to test for a quadratic relationship (b).

galactose concentration alters gene expression only in this plasmid. Figure 2.3.5 shows that at low concentrations of galactose (2g/L) relative fitness does not significantly differ from 1 (wilkoxen signed-rank=-28, p=0.159, N=16). However at higher concentrations, from 5 to 20 g/L, fitness was significantly reduced (relative to 1), with the pRS$^+$ strain having a mean fitness of between 0.836 (s.e.=0.044, N=16) and 0.873 (s.e.=0.033, N=16; table 2.3.6), however a one way ANOVA found no effect significant of galactose concentration on relative fitness across this range of galactose concentrations ($F_{3,62}$=1.275, p=0.291, N=64).

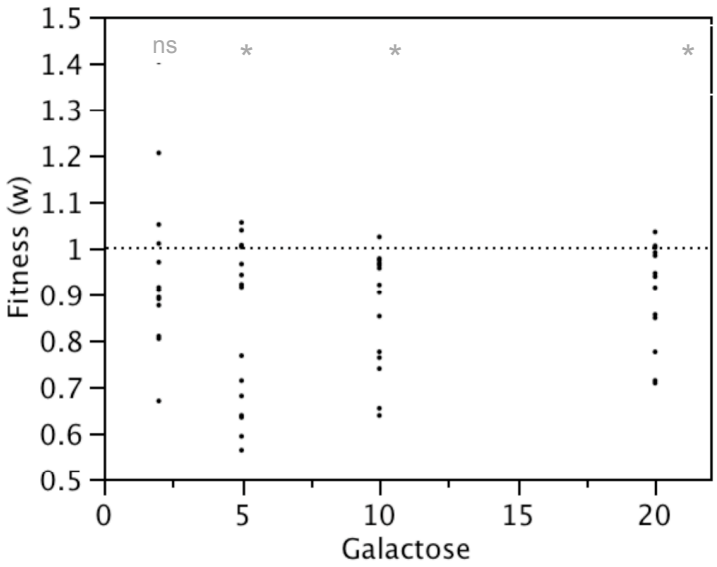


Figure 2.3.5 The effect of galactose concentration on the fitness of [pRS$^+$], compared to [pRS^0] strains. Each point represents one comparison between a [cir^0] [pRS$^+$] strain to an isogenic [pRS^0] strain. The pRS plasmid is a single copy vector-plasmid with the FLP gene under a GAL promoter, therefore increasing galactose concentration is expected to increase Flpp expression but not alter plasmid copy number in the absence of the 2µm plasmid. Significant deviations from 1 (where [pRS$^+$] = [pRS^0] fitness) is shown in grey by ns (not significant) or * (p < 0.05).

Galactose concentration (g/L)	N	Mean fitness (w)	Standard error (w)	Test of difference from 1	
				Signed-Rank	*p*
2	16	0.943	0.046	-28.000	0.159
5	16	0.836	0.044	-56.000	0.0021
10	16	0.881	0.032	-63.000	0.0003
20	16	0.873	0.033	-60.000	0.0008

Table 2.3.6 Summary of relative fitness (w) of strain C [cir^0] [pRS$^+$] compared to strain D [cir^0] [pRS^0] at different galactose concentrations. Treatments tested for difference from 1 using wilkoxen signed-rank test. p<0.0125 (based on the Bonferroni correction) indicates a significant difference in fitness between the pRS$^+$ and pRS^0 strains.

2.4 Discussion

In this study we have investigated the relationship between copy number and fitness in the 2µm plasmid – yeast system. Our results show that by experimentally increasing 2µm copy number, the fitness of the host declines by ~0.09% for every additional plasmid added. In the remainder of this chapter, we discuss (i) potential biases in our estimate of the cost of plasmid carriage, (ii) potential sources of variability in the cost of plasmid carriage that we have not been able to manipulate, and (iii) possible mechanisms underlying the cost of 2µm plasmid carriage.

2.4.1 Slow Growth Phenotype

The results of the control, pRS^0 treatment reveled an unexpectedly high fitness for [cir^+] strains at high concentrations of galactose in the media. In the 20g/L galactose treatment, the [cir^+] strains had an apparent fitness advantage of ≤12% over their [cir^0] equivalent. This finding may represent a genuine benefit to carrying the plasmid, however several studies have found the opposite to be true (Futcher & Cox 1983; Mead et al. 1986), and none to our knowledge have reported an advantage. An alternative explanation for this result is that the strain construction process has introduced a deleterious mutation into the [cir^0] strain. This has been previously reported among strains transformed with a plasmid vector, and defined as the Slow Growth Phenotype (SGP) (Mead et al. 1987). This growth rate reduction can be substantial (>32% in one diploid strain), and appears to be dependent on the presence of extra chromosomal DNA (Danhash et al. 1991). The SGP is not observed in strains transformed by cytoduction, and in transformed strains containing 2µm plasmid DNA (Danhash et al. 1991). However it is not clear whether the fitness cost was ameliorated by the presence of the 2µm specifically, or by extrachromosomal elements in general, and therefore including vector-plasmid DNA. As our transformation and curing process was similar to those described by Mead et al. (1987), we believe this is a likely explanation for the fitness advantage observed in [cir^+] strains. Therefore, the results of the experiment will be discussed in the context of this mutation.

Copy number – fitness relationship

The over-expression of FLP caused an increase in copy number from between 216 – 474 copies from the steady state copy number of ~60 copies per diploid. The resulting fitness cost of 23 – 37% demonstrates the burden this excess of 'parasitic' DNA can have on the host. From this we calculated the per plasmid fitness cost as being ~0.09%. This estimate is likely to be conservative, considering the potential influence of the SGP on

fitness. As fitness is calculated as a relative value of [cir$^+$] compared to [cir^0] growth rate, both having been transformed with the pRS vector-plasmid, we can assume that the SGP mutation is present in both strains. Therefore, under a simple model where the phenotype is expressed in both strains, our estimate of fitness is an accurate reflection of the cost of the 2µm. However, Danhash et al. (1991), found that the SGP was only expressed in the absence of the 2µm, which would imply that only the [cir^0] strain is affected by this mutation. In which case, the fitness cost of the plasmid will be in part countered by the cost of the SGP in the [cir^0] strain, thus making the true cost of the plasmid greater.

2.4.2 Environmental variability in the cost of plasmid carriage

In contrast to the negative relationship between copy number (as a consequence of galactose concentration) and fitness in the pRS$^+$ treatment, the pRS^0 treatment indicated a positive correlation between galactose and fitness. As neither strain in this treatment carries the pRS-GAL-FLP vector-plasmid, copy number was not related to galactose, therefore differences observed are a direct consequence of the galactose sugar concentration, rather than its interaction with the 2µm. This fitness estimate however, will certainly be influenced by the SGP mutation, as the [cir$^+$] [pRS^0] strain did not have a history of transformation, while the [cir^0] [pRS^0] strain has been cured of the 2µm using the pRS-GAL-FLP vector-plasmid, which was itself subsequently removed by counter selection, and can be assumed to carry this mutation. Therefore while the fitness advantage of the [cir$^+$] SGP^0 strain in a high sugar (20g/L galactose) environment was ~12%, in a low sugar environment (2g/L galactose) this strain actually had a ~3% disadvantage. This indicates that there is an additional effect of nutrient availability, which either exacerbates the cost of the 2µm, or ameliorates the cost of the SGP, when galactose concentration is low. Although the former is also possible, the alleviation of the SGP appears to be the best explanation for this relationship. Although counterintuitive, deleterious mutations are often found to be alleviated in stressful environments (Elena & de Visser 2003; Jasnos et al. 2008). In low sugar environments the rate of growth is necessarily slower, covering up the cost of slow growth due to mutation.

Therefore, as we are unable to distinguish between these two hypotheses, we are unable to address whether, and how much the cost of 2µm carriage is altered by environmental variation. In the pRS$^+$ treatment, figure 2.3.3 shows that the high galactose/high copy number treatment is clearly biasing the slope upwards compared to the lower galactose/copy number treatments. The results discussed above suggest that this may be due to environmental interactions with the 2µm or with SGP, however it is also possible that the cost of plasmid carriage diminishes at high copy number. Therefore

again, our estimate of the fitness cost per 2μm plasmid is, if anything, made more conservative by this effect.

2.4.3 Mechanisms for the cost of plasmid copy number

Regardless of this possible underestimation, we have demonstrated that the relationship between copy number and fitness burden is negative. This same pattern has been observed in bacterial-plasmid associations (Patnaik 2000). The mechanism for this cost has two potential sources; through simply carrying and replicating this additional DNA, and the cost of increased protein expression from it. The effect of increasing DNA content has been demonstrated in aneuploid strains of *Saccharomyces cerevisiae* (i.e. strains in which one or more chromosome is duplicated) (Torres et al. 2007). Naturally, the effect of doubling different sets of genes in the genome has very different effects on the host cell, however several features of aneuploidy were universal (Torres et al. 2007). Growth rate was reduced, and the glucose uptake increased, which is consistent with the growth rate advantage of the [cir^0] strains over those containing the plasmid. The effects of aneuploidy were compared against the effect of transformation with YACs (Yeast Artificial Chromosomes), carrying only mammalian genes, which are not expressed in the yeast cell (Torres et al. 2007). Although some metabolic responses were elicited, in general YACs did not cause the drop in growth rate seen among aneuploid strains. One exception to this was the largest YAC insert of 1.6Mb, which showed a short delay in growth rate delay (5 minutes) compared to between ~25-40 minutes for similar sized chromosomal DNAs. This amount of DNA is equivalent to 250 2μm plasmid genomes and therefore at the low end of plasmid content in [pRS^+] strains. Therefore, the cost of additional DNA in the cell can be in part attributed to the DNA itself. Considering that 1.6Mb of DNA represents >10% increase in total cellular DNA, this is not surprising.

However the cost of DNA itself is hugely outweighed by the cost of gene expression, which requires transcription, translation, and also an increase in the activity of the protein degradation pathway to counteract the rise in protein levels (Torres et al. 2007). In this study we have measured the cost of gene expression from the pRS-GAL-FLP vector-plasmid, in which the FLP gene was being over expressed. Fitness was reduced by ~15% as a consequence of the higher levels of galactose induced gene expression. This represents a huge reduction in fitness caused by the over-expression of just one gene from a single copy plasmid. More recently, a study by Stoebel et al. (2008) isolated the cost of expression in the Lac operon in *E. coli* to the occupation of ribosomes and the use of RNA polymerase, therefore suggesting that the cost is largely due to use of cellular facilities, rather than raw materials *per se.*

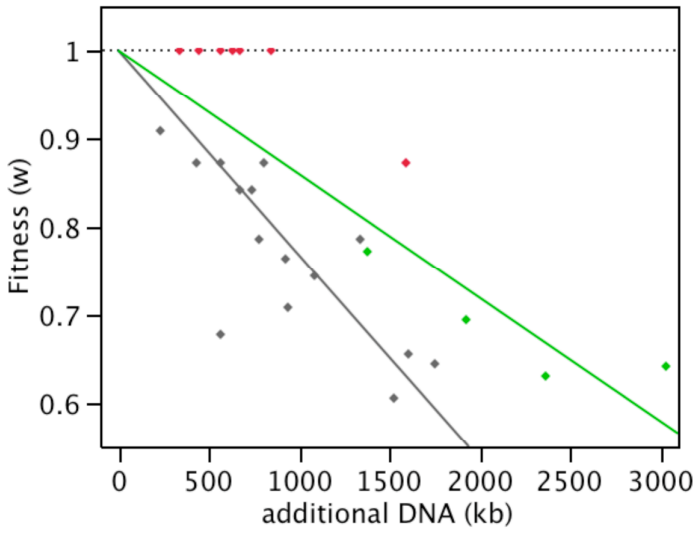


Figure 2.4.1 Comparison of the fitness cost of additional DNA from chromosomes (grey), YACs (red) and 2µm plasmids (green). Data for chromosomal and YAC DNA was taken from Torres et al. (2007) and 2µm data shows the mean fitness for each galactose/copy number treatment (this study). Line fitted through chromosomal data (grey) shows a slope of -0.000232 (s.e.= 1.639 x 10^{-5}, t= -14.14, p<.0001), compared to an estimated slope of -0.00014 (s.e.= 1.189 x 10^{-5}, t= -11.79, p=0.0013) in plasmid DNA (green).

Figure 2.4.1 shows a comparison between the results of Torres et al. (2007) and our own. Slopes taken from chromosomal DNA shows a predicted cost of 0.023% per kb of extra DNA in the cell. In comparison, plasmid DNA was slightly 'cheaper', at a cost of ~0.014% per kb (or 0.09% per 6.4kb plasmid). This difference may reflect the bias' discussed earlier in this chapter, however it is also possible selfish genetic elements such as the 2µm plasmid would be cheaper than chromosomal DNA. Although all DNA is to some extent, subject to selection for molecular efficiency (Akashi 2003), this selection may be expected to be stronger in selfish genetic elements. In particular, in elements to which there is a fitness advantage in being numerous, such as the 2µm, selection would favour cheaper elements that can reach higher copy numbers whilst imposing the same fitness cost on the host. Equally, in high copy number elements, the selective advantage of adaptations which reduce the cost per copy of an element is amplified by the elements copy number. Consequently, small difference in the cost of individual elements may be considerable if these elements are at high copy number. For instance a ~0.001% fitness advantage of a mutation in one copy of an element becomes a 0.1% fitness advantage when there are 100 copies. Therefore, we may expect to see adaptations such as smaller proteins, tight gene expression control (Akashi 2003) and high codon usage bias (Tuller et al. 2010) which can maximise the efficiency of protein transcription and

translation in order to minimize the cost of each element. Analysis of the genomes of three different variants of the 2μm in chapter 4 show that all plasmid genes do in fact have exceptionally high codon usage bias. Each of the four genes fall in the top 1% of a survey of codon bias in >3000 nuclear genes (chapter 4).

2.5 Conclusions

The estimates of the fitness cost of plasmid burden presented here are likely to be conservative. Despite this, the cost of carrying the plasmid at high copy numbers was extremely high. At the highest copy number of 475 (~8x the 'normal' copy number for this strain) the plasmid imposed a 37% reduction in fitness. The implications of this increasing fitness cost are great when considering the evolution of the plasmid as a selfish element in the yeast cell. Increases in copy number will incur a cost, and therefore be under negative selection from the host. This suggests that high copy numbers are being driven by positive selection acting at the plasmid level. The possible sources of this positive selection are addressed further in chapter 3. Briefly, these include reducing the rate of plasmid loss through miss-segregation, and gaining a replication advantage during competition within the host.

2.5.1 Variability in the Cost of Plasmid Carriage

We have placed an estimate on increasing the cost of plasmid carriage, however it is unlikely that this cost is universal among plasmids and strains, or indeed fixed over the course of plasmid-strain interactions. Plasmid copy number varies between 2 and 753 in *S. cerevisiae* and *S. paradoxus* strains (Chapter 4), which would imply that the burden imposed by 2μm plasmids can reduced host fitness by up to 68% (based on the estimated fitness cost of 0.009% per plasmid) and still be maintained in the population. As selfish genetic elements are theoretically able to impose a maximum of 50% fitness cost to hosts in a completely outcrossing population (Futcher et al. 1988), this level of cost is highly improbable. In reality, the cost of plasmid carriage is likely to vary by genotype, and possibly also by environment.

Among bacterial – plasmid associations the cost of plasmid carriage is generally ameliorated over time in evolutionary experiments. Different responses are seen in different plasmid systems and range from a general increase in tolerance by the host (Bouma & Lenski 1988; Modi & Adams 1991), general adaptations in both host and plasmid (Dahlberg & Chao 2003; Dionisio et al. 2005), and host-specific plasmid adaptation (Modi & Adams 1991). In the 2μm – yeast system the large variation in copy number implies that genotypic variation in the cost of this association will also be large.

We have argued that the majority of the cost of 2μm carriage is in the transcription and translation of addition protein. Therefore, an obvious target for selection is the molecular efficiency of the nuclear and plasmid genomes, which could be increased by reducing gene length or expression (Akashi 2003). In addition, we may also expect the cost of plasmid carriage to be resource dependant. Low quality environments have the potential to be either ameliorating, as is the case for many deleterious mutations (Elena & de Visser 2003; Jasnos et al. 2008), or exacerbating, by increasing the competition between host and nuclear gene expression. Therefore we would expect that the 'price' of plasmid copy number is a consequence of all these factors, and that high copy number strains will be under stronger selection for a reduction in the cost of each plasmid genome.

2.5.2 Future Directions

Although we do not expect the impact of the SGP would negate the main conclusions of this study, its presence confounds our understanding of the impact that galactose treatment has on the outcome. However, it is possible that the true cost of the 2μm is higher than predicted. To assess this hypothesis, an alternative protocol using cytoduction to transform strains with the pRS-GAL-FLP plasmid should be used.

We have also not addressed the consequence of reducing plasmid copy number. The predictions of this study would indicate that fitness costs will be lower with a smaller plasmid burden. Although this is supported by the fitness advantage observed in [cir^0] strains compared to their [cir^+] counterparts, we cannot fully reject the idea that plasmid steady state copy number is an optimum for both host and plasmid. This could potentially be circumvented by using 2μm plasmid in which the native FLP is knocked out, leaving copy number solely under the control of the pRS-GAL-FLP vector.

In addition, to the improvements on this protocol an important extension to this work is to determine the variability of this cost. As we have described, the existing variation in copy number among strains of *S. cerevisiae* and *S. paradoxus* strongly suggest that the cost of plasmid carriage is not fixed, but dependant on both host and plasmid genotypes, and possibly also environmental factors. This could be tested by quantifying variability in the fitness cost of different plasmid genotypes in different host backgrounds under rich and poor environmental conditions. A basic prediction of this experiment would be that cost of each plasmid is inversely related to its copy number in the host.

Chapter three

Sex and Selfishness: The Influence of Sexual Strategy on Virulence Evolution in the 2 Micron Plasmid of *Saccharomyces cerevisiae*

3.1 Introduction

3.1.1 Sex and intragenomic conflict

Sexual reproduction is one of the great innovations in the evolutionary history of life. However, for all its advantages, sex has also provided the basis for the evolution of selfish genetic elements, a diverse group of genetic units which parasitise the genome for their own replication. Under asexual reproduction, the whole genome is inherited as a single unit, consequently, each allele shares the same evolutionary fate as the whole of the genome, and by extension, the individual. Therefore selection acting on the individual favours a 'harmonious' genome, which maximises the fitness of the individual. In contrast, under sexual reproduction, the association between alleles is transitory. Through genetic recombination the linkage disequilibrium within the genome is broken down and alleles are shuffled into novel backgrounds. As a consequence, alleles, or linked groups of alleles, become the smallest unit of selection (Hurst 1992). Where selection on the gene and selection on the individual do not favour the same out-come, intragenomic conflicts can emerge (Partridge & Hurst 1998).

The evolution of these conflicts is dependant on a difference in transmission potential among the genes or genetic elements. Elements able to 'drive' (transmit to >50% of offspring; Burt and Trivers 2006), have a distinct advantage over those confined by the laws of Mendelian inheritance, as they transmit to more meiotic progeny and can therefore increase in frequency in the population with out relying on positive selection (Hickey 1982). Clearly, as this transmission advantage is dependant on meiosis, this is only possible during sexual reproduction.

In agreement with this, the spread of various vertically transmitted selfish genetic elements has been shown experimentally to be dependant on sex. The homing endonuclease gene VDE (Goddard et al. 2001), a retrotransposon Ty-3 (Zeyl et al. 1996) and the 2μm plasmid (Futcher & Cox 1983) have all been found to spread in experimental sexual *S. cerevisiae* populations, but failed to do so or declined under asexual reproduction. In *Daphnia pulex* transposon abundance has been found to be significantly higher in sexual isolates than in obligate asexuals (Schaack et al. 2010; Valizadeh & Crease 2008). Likewise, among the asexual Bdelloid rotifers, vertically transmitting selfish genetic elements are conspicuous by their absence. Among the five species in this taxa, no LTR, or non-LTR, retrotransposons were identified despite being found in all species of the sexual sister class, *Monogononta*, and more basal taxa indicating that they have been lost subsequent to the loss of sex (Arkhipova & Meselson 2000). They do however contain DNA transposons, which have a greater ability to transmit horizontally (Arkhipova 2005).

3.1.2 Sex and selfish genetic element virulence

If sex allows the spread of these selfish elements, the amount of sex is expected to influence the extent of their selfishness. Obligate asexual and sexual organisms represent two ends of a continuum, with facultatively sexual and selfing organisms in between (Bestor 1999; Hurst & Werren 2001). While purely asexual reproduction may preclude the evolution of intragenomic conflict, only very small amounts of sex is required to break down linkage between alleles (Charlesworth et al. 1993), therefore even small amounts of sexual reproduction can generate conflict. In addition to the frequency of sex, the rate of inbreeding vs outcrossing is also important in determining the rate of spread though the population, as the benefits of drive are mitigated in highly homozygous populations. Consequentially, elements cannot invade inbred populations, which has been demonstrated experimentally in the 2μm plasmid (Futcher et al. 1988) and VDE (Goddard et al. 2001) of *S. cerevisiae*. Therefore, the degree and rate of outcrossing in a population sets the tempo of selfish genetic element transmission.

We can expect that the virulence (i.e. the fitness cost) of the host-selfish genetic element association will evolve to maximise transmission, as a consequence of a tradeoff between exploitation and harm (Bull 1994; Frank 1996). On one side, transmission may be increased by higher levels of host exploitation by the element. However this is balanced by the increasing cost of reducing the fitness of the individual, a cost which is also being borne by the element. The rate of sex, and therefore transmission, plays a key role in determining the strength of this negative selection, as it determines the time spent in association with a single genome, and therefore, the extent to which it suffers from its own fitness burden

(Bestor 1999; Hurst & Werren 2001). Consequently, the amount of sex shapes the outcome of this equilibrium, strongly selecting against virulence when outcrossed sex is rare.

However, this optimization of transmission assumes high relatedness among elements within the individual (Frank 1996). Where only one genotype inhabits an individual, this tradeoff is the main determinate of fitness. More costly elements are removed from the population by selection on the individual. As the number of unrelated genotypes increases, competition between them will drive virulence higher, beyond the level at which transmission is optimised (Nowak & May 1994). By over exploiting the individual an element may gain a transmission advantage over a rival, for instance by being passed to a higher percentage of progeny. Therefore, a mixed infection is analogous to a 'tragedy of the commons' where genotypes maximise their own replication relative to rival elements within the individual, but in doing so reduce the reproductive rate of the individual and therefore the group (Buckling & Brockhurst 2008). The rate of sex, and degree of outcrossing will alter the level of relatedness by greatly increasing the probability of different genotypes being combined into one individual. However, it is also possible for competition between unrelated elements to be internalized, so that elements compete by repressing the replication of the other, rather than competing for host resources. Therefore, in this way, competition between unrelated elements can reduce cost of selfish genetic element burden to the individual (Turner & Chao 1998).

3.1.3 Aims

The goal of this chapter is to test the hypothesis that the sex drives increasing levels of virulence in selfish genetic elements. This will be a consequence of increasing the rate of spread through the population, which reduces the strength of selection against virulence, and decreases the level of relatedness among elements within an individual.

3.1.4 Experimental system

Saccharomyces cerevisiae

The bakers yeast, *Saccharomyces cerevisiae* represents an ideal model organism for the study of the evolutionary impacts of sexual reproduction. As a microbial model system it can be grown rapidly in large populations and cultures can be kept indefinitely at –80^{o}C and revived at a later date for analysis. It also has a facultatively sexual reproductive system (*see* figure 1.1, Chapter 1), in which meiosis can be induced by starving cells of nitrogen. In order to control for this environmentally induced sexuality, we have used a strain developed by Goddard *et al* (2005) which is capable of undergoing the sporulation process, but is

prevented from going through meiosis by the interruption of two genes; *spo11,* which prevents meiotic recombination, and *spo13,* which prevents reduction division (Goddard et al, 2005). Following sporulation, therefore, this strain produces an ascus containing two identical diploid spores rather than the normal four haploids, and consequently cannot undergo sexual reproduction. In addition, we have used an alternating marker selection technique, which selects for homozygotes and then heterozygotes at two marked loci in each transfer, therefore ensuring complete and consistent outcrossing in sexual treatments. We believe that this system provides a powerful protocol for testing the effect of sexual reproduction in long-term experiments.

The 2 Micron plasmid

The 2μm is an endogenous high copy number plasmid found in strains of *S. cerevisiae* and its sister taxa, *S. paradoxus* (chapter 4, this thesis). Copy numbers are extremely variable between strains, ranging from 2 to 753 copies per diploid cell (chapter 4). All plasmid gene products are involved in its own replication and persistence, and therefore, while it confers no benefit to the host, it does not actively cause harm to the genome either. As shown in the Chapter 2, increases in copy number have a negative relationship with fitness of the cell, with an estimated 0.09% fitness cost per additional plasmid. Copy number therefore represents the primary source of virulence in this element, as increasing numbers of plasmids mean increasing amounts of proteins to be produced which are not beneficial to the cell (reviewed in Chapter 2).

Plasmid copy number is controlled by both the plasmid itself and the host genome. All four plasmid proteins are involved in a self-regulating feed back loop, both to increase copy number when it dips, and to inhibit replication when it is high (Murray et al. 1987). However this system is also dependent on the host genome. Mutations causing malfunctions in the cell's protein degradation pathway lead to copy number increases from 2 to 68 times that of the wild type and must therefore play a key role in suppressing plasmid over replication (Burgess et al. 2007; Chen et al. 2005; Dobson et al. 2005; Holm 1982; Sleep et al. 2001; Xu et al. 1999). Considering the fitness cost of around 0.09% per additional plasmid, and the fact that the plasmid actively regulates its own copy number at a given threshold, it is clear that high copy number is adaptively significant to the plasmid.

We have used this yeast-plasmid system to test the hypothesis that sexual reproduction permits the evolution of higher virulence of a selfish genetic element in experimentally evolving populations. This study represents the first empirical test of this principle to our knowledge. We can expect that under asexual reproduction the dominance of individual

level selection will maintain a more conservative element. Conversely we anticipate that sex should allow more costly elements, with elevated copy number to evolve. We also decided to test for the presence of another potential selfish genetic element, Selfish Mitochondria. Mitochondria frequently develop loss of function mutations, which occasionally also show an ability to drive by out-replicating normal mitochondria in the colonization of new cells (Taylor et al. 2002). We therefore expected that sexually reproducing populations would also allow for the invasion of these mutants in addition to changes in the 2μm plasmid.

3.2 Materials and methods

3.2.1 Strains and culture

The sexual strain S, was produced by a cross between haploid strains DH89 *MATα ho lys2Δ* and DH89 *MATa ho ura3Δ* (Goddard et al. 2001). The resulting diploid is therefore heterozygous at the URA3 and LYS2 loci: DH89 *ho/ho a/α URA3/ura3Δ lys2Δ/LYS2*.

To control for the sexual treatment, an obligate asexual strain, A, was also used. MG98 (*a/α ho/ho ura3Δ/ura3Δ spo11Δ/spo11Δ spo13Δ::KanMX4/spo13Δ::KanMX4)* has the genes *SPO11* and *SPO13* deleted, preventing cells from undergoing meiotic recombination and reduction division respectively. Therefore it is able to undergo a form of sporulation, resulting in an ascus containing two identical diploid 'spores' (Goddard et al. 2005).

The experiment was divided into 3 treatments, Sexual, Asexual and Budding (described in table 3.2.1) with 8 replicate populations per treatment. The two ancestral strains were plated onto Minimal Media (MM) agar (reagents are described for all medias and solutions in appendix I) for strain S, or MM + uracil for strain A (a uracil auxotroph). Eight individual colonies were picked at random from the relevant strain and grown to saturation at 30^{o}C in 10ml MM or MM + uracil for strains S and A respectively, with shaking (100rpm). Cultures were washed in 0.9% saline and half of the culture removed

Treatment	Purpose	n	Strain	Procedure
Sexual	Experimental	8	S	Budding growth followed by sporulation and randomized mating
Asexual	Treatment control	8	A	
Budding	Strain control	8	S	Budding growth only

Table 3.2.1 Three treatments were used in the experiment. Two control treatments were included to control for the effect of the protocol and also for the effect of the strain.

for storage at -80^{o}C. For the Budding treatment the remaining culture was resuspended in sterile saline and incubated at 10^{o}C. In the Sexual and Asexual treatments the culture was resuspended in 20ml sporulation solution and incubated for 6 days at 30^{o}C, 100rpm, to induce sporulation. Following this, cultures were digested in 3ml outcrossing solution at 37^{o}C for 3hrs to remove the asci wall and loosen the intra-tetrad bridges between spores in the tetrad. They were then resuspended in 10ml sterile saline and sonicated at 40amps, for 100secs in 2 sec pulses to disrupt tetrads and randomize spores. The resulting spores were resuspended in 1ml YPD and left static for 3hrs to encourage germination and mating. This process makes up one 'transfer'. Following each mating, the culture was resuspended in 1ml sterile water, and half (0.5ml) was removed for freezer stocks. To this, 0.5ml 60% glycerol solution was added (to a final concentration of 30% glycerol), mixed by inversion and stored at -80^{o}C. The remaining cells were then carried on to the following transfer by washing cultures in 10ml sterile saline before resuspension in 10ml MM (+/- amino acid supplementation as described below). A total of 22 transfers were completed.

Total asexual generations per transfer are estimated as approximately 5.463, using the equation $g = (\log_{10}(N_t) - \log_{10}(N_0) / \log_{10}(2)$, where g is the number of generations, N_0 is the population size at time 0 and N_t is the population size at time t. At saturation cultures reach a cell density of approximately 1.5 x 10^7 cells/ml. Therefore population size reaches a maximum of roughly 1.5 x 10^8 cells in 10ml of media. Based on the average observed sporulation rate of ~10% during the experiment (data not shown) and the assumption of fully randomised mating, in which only ~45% of progeny are able to grow due to the differential selection described below, cultures were bottlenecked to approximately 3.4 x 10^6 cells.

Amino acid supplementation

The heterozygosity of strain S at the URA3 and LYS2 loci was used to ensure that sex was maintained throughout the experiment. For odd transfers the cultures were grown in MM without amino acid supplementation, therefore selecting for the URA+ and LYS+ wild type phenotype. Following even transfers homozygous mutants of the 2 auxotrophies were selected separately by splitting the culture in two and resuspending half in MM + uracil + 5FOA, selecting for ura- LYS+ phenotypes, and the other half in MM + lysine + αAA, selecting for URA+ lys- phenotypes. Consequentially, phenotypes selected for in one transfer are unable to grow in the next, therefore selecting for mated diploids only (this is further described in figure 3.2.1).

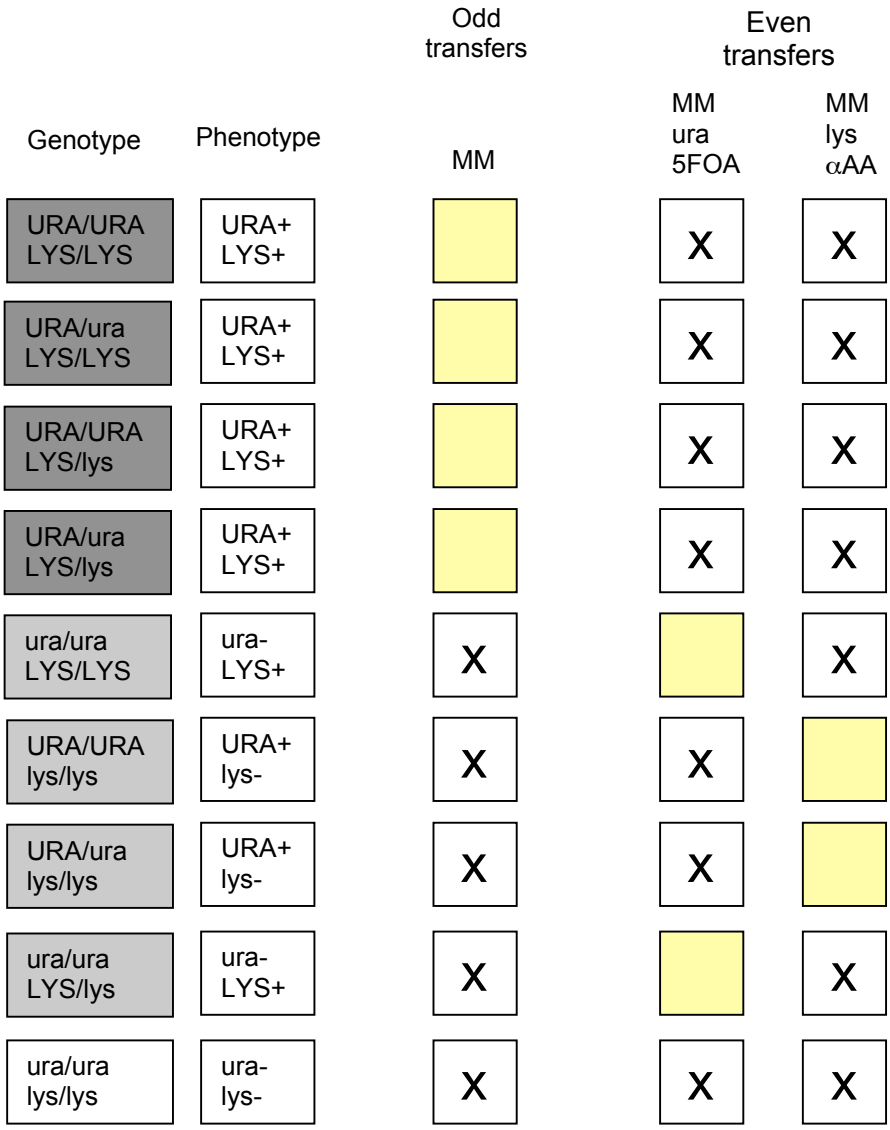


Figure 3.2.1. Media alternation method used to select for mated cells in the Sexual treatment. Genotypes able to grow (yellow) in odd transfers are unable to grow (cross) in odd transfers, and vice versa. As the only way switch between the phenotypes being selected is to reproduce sexually, sex is actively selected for.

To control for the effect of varying chemical environment in the asexual treatment, the media alternated between MM + uracil for odd transfers and MM + uracil + 5FOA for even. Budding populations were always grown in MM without supplementation.

3.2.2 Analysis of evolved populations: Population level analysis

2 Micron Plasmid copy number

To estimate overall change in plasmid virulence, DNA was extracted from 200µl aliquots from freezer stocks of populations at transfers 0, 10 and 20. As half the population was stored in 1ml of glycerol solution, this represents 10% of the total population following mating.

Cultures were washed in water and the DNA extracted using the Promega Wizard® SV 96 Genomic DNA Purification System as described in appendix I. DNA was also extracted from 200µl cultures of ancestral strains S and A, their progenitor strain Y55 and a plasmid free control strain YDG 598 [cir^0] (*see* chapter 2).

DNA extractions were diluted 10 fold and amplified with primer pairs targeting the HO gene as a yeast genomic target (qHO_for CGGTTTGTGACGATCACGTT; qHO_rev CCTGTTTTCATCTGGACCATCTC and the FLP gene of the 2µm plasmid (qFLP_for TTCAATGAAGGGCCTAACGG and qFLP_rev TCGCTCCAATTTCCCACAAC). To check the efficacy of these primers, and to ensure that sequence mutation or genome rearrangements in the evolved populations have not influenced the binding of these primers we compared copy number estimates for sexual populations from transfers 0 and 20 using alternative plasmid targets; the inverted repeat regions (IR for = AGGTGCGACGTGAACAGTGA and rev = GAAAATGCAACGCGAGCTG) and the STB region (STB for = CATCCCCGGTTCATTTTCTG and rev = AGATATGCTATTGAAGTGCAAGATGG) Comparisons between primers are shown in figure 3.2.2.

To control for any stochastic effects of well position, plate etc, qPCR plates were randomised. Populations 1-8 from each treatment were randomly split into two blocks, therefore each block contained 4 populations of each treatment at the 3 time points. Within each block, well positions were randomised with the exception of keeping the qHO and qFLP reactions for each sample in adjacent wells. In addition to samples, each plate included the control strains, no-template control and a standard curve of using strain Y55 (concentrations, 100, 10, 1, 0.1 and 0.01ng/reaction). Each sample was run once per plate and each plate repeated 3 times. Reaction conditions and the calculation of copy numbers were as described in appendix I. Copy number was calculated independently for each replica PCR and these estimates averaged.

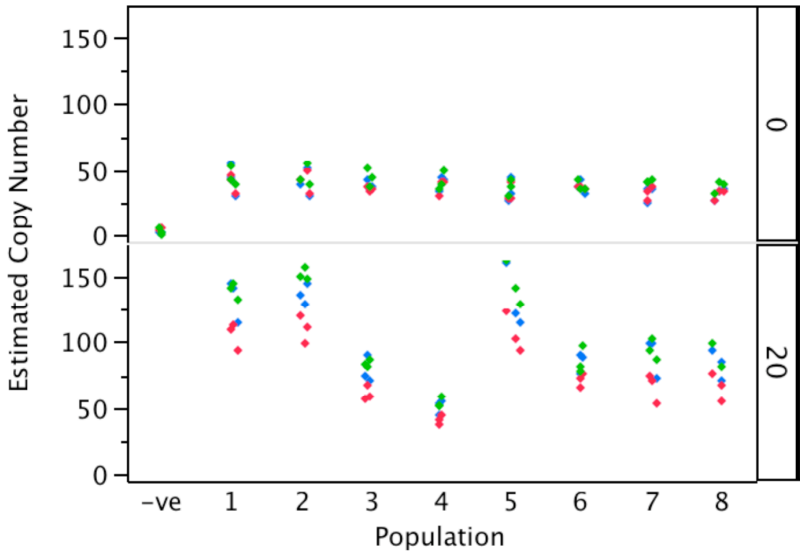


Figure 3.2.2 Comparison between copy number estimates based on primers targeting FLP (blue), STB (green) and the inverted repeat regions (red) at transfers 0 and 20 in sexual populations, and the negative control YDG598 [cir^0] (-ve). Each point represents an individual estimate of copy number (replicates = 3).

Plasmid genome rearrangements

To check for major rearrangements in the plasmid genome we amplified 1-2kb sections around the plasmid from sexual populations at transfers 0 and 20, and run on a gel to determine any differences in amplicon length. It is possible that plasmids in which sections of the genome are deleted would have a replication advantage, as shorter genomes can be replicated faster. As all plasmid gene products are shared between plasmid genomes, loss of function in one or more genes would allow such mutants to be sustained in mixed populations. Smaller plasmids may conceivably have a replication advantage over the wild type, as they could be replicated faster and therefore to higher copy number. PCR fragment coverage is shown in figure 3.2.3. PCR and gels were performed as described in appendix I.

Screening for mitochondrial mutants

Populations were screened at transfers 0, 5, 10, 15, 20 and 22. Cultures were plated onto YPD agar to give 3 plates of ~200 colony forming units (CFUs) per plate. These were grown at 30°C for 2 days and replica plated onto YPG, which contains glycerol as the sole carbon source (Amberg et al. 2005). Glycerol is non-fermentable and therefore screens for the ability to respire. These were grown at 30°C for 2 days. Colonies were then scored for the ability to grow on YPG.

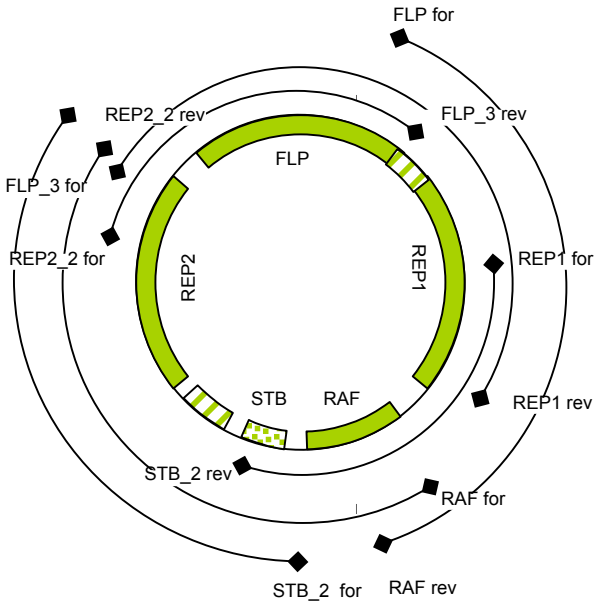


Figure 4.2.3 PCR amplification of the 2µm to check for genome rearrangements. Each section of the plasmid was amplified at least twice using difference primer pairs as shown. Diamonds represent primers, and black lines, amplified regions. Primer sequences are shown in appendix I.

Fitness

Growth rates were measured in individual colonies rather than from whole population cultures to avoid the complication of variants out competing each other during the assay. 50-200 cells were plated onto MM + uracil + lysine agar and 15 colonies picked using sterile toothpicks. Colony choice was randomised using a grid overlaid onto the plates and the colony closest to each the intersection, measured from the center of the colony to avoid size bias, was picked. These were used to inoculate 200µl MM + uracil + lysine and left to grow for 48hrs. 10µl of these cultures were then diluted into a further 200µl of media in 96 well optical plates. A total of 10 plates were used, each containing 2 colonies from each population at each time point, with well positions randomised. Growing cultures were read every 2hrs using a Biotek Synergy HT plate reader at OD 600nm for 60hrs and stored at 30^{o}C between reads.

Growth rate was estimated using plate reader software, Gen5, which calculates the maximum linear slope, in mOD minute^{-1}, for each growth curve by sequentially calculating the slope across 5 adjacent 2hr time points and returning the maximum value (maxV). As turbidity is proportional to log cell density we can use the change in OD 600nm absorption as a measure of population growth (µ) in the equation $N_t = N_0 e^{\mu t}$. A relative fitness estimate of transfer 20 compared to transfer 0 can then be calculated from the equation (Hartl & Clark 1997):

$$w = e^{\left[\frac{\mu_{20}}{\mu_0} - 1\right]}$$

Cells from saturated growth plates were patched onto YPG agar using a 96-pin replicator and grown for 48hrs at 30^{o}C to identify mitochondrial mutants. Cultures which failed to grow were reconfirmed by plating 5µl of the culture onto YPD and YPG agar, to distinguish those which did not grow at all from those which could not respire.

3.2.2 Analysis of evolved populations: *Within Population Variation*

Plasmid Copy Number

To gain an estimate of copy number variation, and to check for the possibility of spontaneous plasmid free cells arising, which would skew estimated copy number downwards, colonies picked from the growth rate assay were also measured for 2µm copy number. Following growth to saturation in which growth rate was measured, cultures were washed and DNA extracted as before in the same 96 well format. Quantitative PCR was conducted as before with each reaction repeated twice.

Fitness

Colony level fitness was calculated for each colony from transfer 20 populations, using the equation shown above where μ_0 = mean maxV of all colonies at transfer 0.

3.2.3 Analysis

Statistical analysis was performed as described in the text using JMP 8 (SAS Institute Inc, Cary, NC, USA). Fully factorial models, including random effects of plate (PCR or growth plate replicates) and block (within plate layout) were fitted and then simplified by removing non-significant factors, until the minimum model was reached, using a threshold of $p<0.05$. Where multiple comparisons were used for further analysis, the alpha level was divided by the number of analyses performed, according to the Bonferroni correction.

Covariation between dependant variables was analysed using Reduced Major Axis (RMA) regression (Sokal & Rohlf 1981). Correlations were performed using Pearson Product Moment correlation to test for a significant covariance between 2 factors. Where $p<0.05$, an orthogonal regression was performed using the standard deviation of the Y variable divided by the standard deviation of the X variable to determine an error variance ratio for each line.

3.3 Results

3.3.1 Copy number evolution

All ancestral populations contained approximately 40 plasmids per cell at the start of the experiment (mean=40.828, s.e.=1.376, N=24), with no differences between treatments (ANOVA Treatment: $F_{2,21}$=1.147, p=0.337, N=24, table 3.3.2). Copy number change was analysed using repeated measures, and we identified a significant treatment by transfer interaction (repeated measures MANOVA treatment x transfer: $F_{4,40}$=11.048, p<.0001, N=24; table 3.3.1). In the sexual treatment copy number increased in all but one population after 20 transfers (repeated measures MANOVA Transfer(sexual): $F_{2,6}$=18.125, p=0.0029, N=8; α-level adjusted to 0.0167 following Bonferroni correction, table 3.3.3). Figure 3.3.1 shows that this increase is not linear, but that copy number increases in all populations from transfer 0 to 10 (mean change in copy number =+15.997, s.e.= 2.485, N=8). Between transfers 10 and 20 however, half of the populations continued to increase in copy number, with three reaching over 3 times their ancestral copy number, while four populations declined to near ancestral copy numbers. Both asexual and budding treatments decreased on average over the 20 transfers (mean change in copy number =-6.699, s.e.=3.850, N=8 and mean change in copy number =-10.572, s.e.=2.947, N=8 respectively). The effect of transfer was significant in the asexual treatment but not in the budding treatment after correcting for multiple comparisons (repeated measures MANOVA Transfer(asexual): $F_{2,6}$=15.828, p=0.0040, N=8; and Transfer(budding): $F_{2,6}$=5.559, p=0.0431, N=8; α-level adjusted to 0.0167 following Bonferroni correction, table 3.3.3.

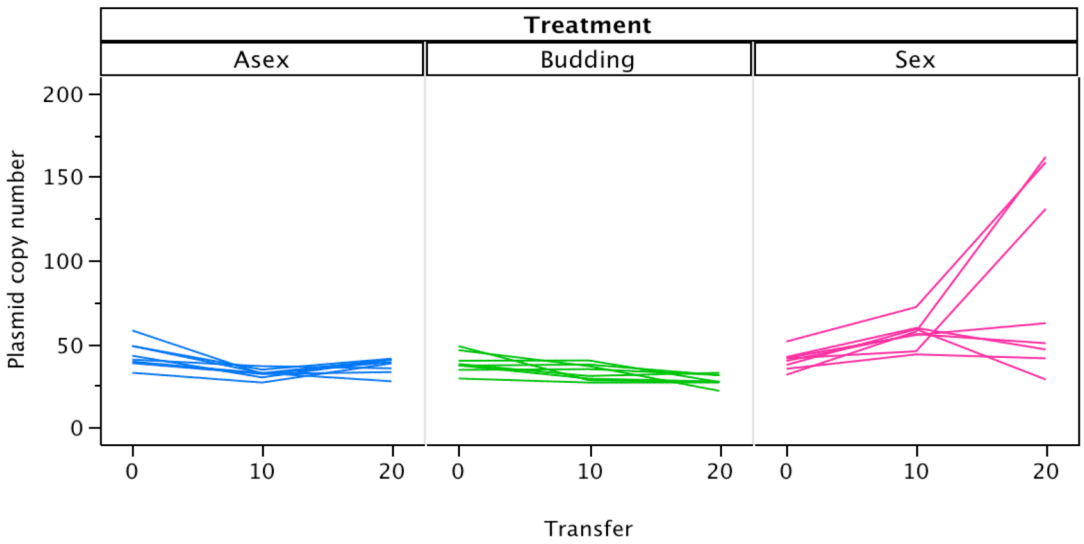


Figure 3.3.1 Change in 2µm plasmid copy number over 20 transfer cycles in reproductive treatments, sex (pink), budding (green) and asexual (blue). Each line represents the mean estimate for each population (N=3).

Source	Num df	Den df	*F*	*p*
Between subjects				
Treatment	2	21	11.821	0.0004*
Within Subjects				
Transfer	2	20	1.170	0.331
Transfer* Treatment	4	42	7.1447	0.0002*
N=24				

Tables 3.3.1 Repeated measures MANOVA of 2 micron copy number measured after 0, 10 and 20 transfer cycles under different reproductive treatments.

Transfer	Source	df	Mean sq	*F*	*p*
0	Treatment	2	51.482	1.147	0.337
	Error	21	44.8959		
10	Treatment	2	1462.347	40.110	<.0001*
	Error	21	36.46		
20	Treatment	2	7484.889	7.181	0.0042*
	Error	21	1042.31		
N=24					

Tables 3.3.2 One sample ANOVA of treatment at each time point. Tukeys HSD comparisons showed that in transfers 10 and 20 the sexual treatment had a significantly higher copy number than the two asexual treatments, which showed no difference from each other. Significance level adjusted to p=0.0167 following Bonferroni correction.

Treatment	Source	Num df	Den df	*F*	*p*
Sexual	Intercept	1	7	70.610	<.0001*
	Transfer	2	6	18.125	0.0029*
Asexual	Intercept	1	7	1455.120	<.0001*
	Transfer	2	6	15.828	0.0040*
Budding	Intercept	1	7	1234.499	<.0001*
	Transfer	2	6	5.559	0.0431
N=24					

Tables 3.3.3 Repeated measures MANOVA testing the effect of transfer individually for each treatment. Significance level adjusted to p=0.0167 following Bonferroni correction.

Copy number was also measured in individual colonies at transfers 0 and 20. These estimates varied widely within every population at both transfers, and no cultures measured appeared to have lost the plasmid. 59 of the 720 colonies failed to grow during the experiment (defined as having a maximum OD600nm of <0.2) and were therefore excluded from the analysis. Colony level measurements of copy number (square root transformed to normalize residuals) were analysed using ANOVA. There was a significant effect of population (ANOVA population[treatment]: $F_{2,634}$=1.659, p=0.0326, N=660: table 3.3.4) and a significant interaction between treatment and transfer (ANOVA treatment x transfer: $F_{2,634}$=9.327, p<.0001, N=660), as mean copy number increased in five of the eight sexual populations but remained low or decreased in the asexual populations. Copy number varied widely between colonies within each population. In the sexual treatment at transfer 0, for example, copy number ranged between 19.419 to 123.127 copies per cell. This variation was consistent across all populations, with no significant effect of treatment, population or transfer, tested using ANOVA on standard deviation of copy number for each populations (ANOVA treatment: F=0.436, p=0.543, N=46, ANOVA populations[treatment]: F=0.721, p=0.770, N=46, ANOVA transfer: F=1.007, p=0.327, N=46, table 3.3.5). Figure 3.3.2 shows the means estimated from the colony level assay, compared to estimates from population level extractions.

Source	df	Mean Sq	*F*	*p*
Treatment	2	1.518	0.790	0.4541
Population[Treatment]	21	3.187	1.659	0.0326*
Transfer	1	0.400	0.208	0.6484
Treatment*Transfer	2	17.917	9.327	0.0001*
Error	634	1.921		
			***F*=4.032, *p*<.0001, *N*=660**	

Table 3.3.4 ANOVA on estimated copy number (square root transformed). 15 colonies of each population were sampled from transfer 0 and 20 populations. Cultures with maximum OD 600nm reading <0.2 were deemed not to have grown and therefore excluded for the analysis.

Source	df	Mean Sq	*F*	*p*
Treatment	2	73.570	0.863	0.436
Population[Treatment]	21	61.477	0.721	0.770
Transfer	1	85.784	1.007	0.327
Error	21	85.224		
			***F*=0.960, *p*=0.543, N=46**	

Table 3.3.5 ANOVA on standard deviation of estimated copy number measured in 15 colonies from each population sampled at transfer 0 and 20. Several cultures did not grow, and were therefore excluded from analysis. In two populations no estimate of variance was possible due to lack of viable cultures.

There was no significant difference between means estimated from the two different methods (matched pairs t-test: t_{46}=-1.335, p=0.1884, N=47), however the three high copy number populations seen at transfer 20 showed considerably lower colony level means (figure 3.3.2). Furthermore, in populations 2 and 5, the population level estimate exceeded the maximum copy number measured at the colony level. Conversely, the colony level mean of population 4 in the sexual treatment was much higher than the population level mean.

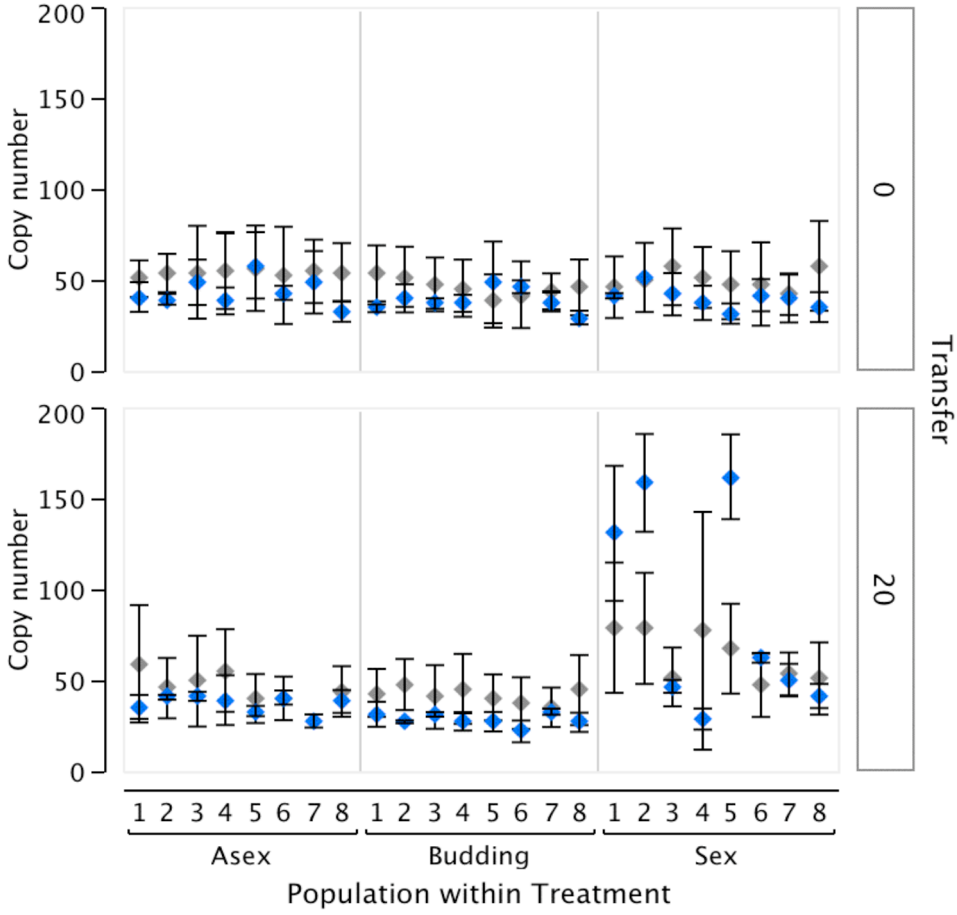


Figure 3.3.2 Comparison of 2µm plasmid copy number estimates produced from qPCR on whole population DNA extractions (blue) and from 15 randomly sampled colonies from each population (grey). Error bars represent standard deviation from the mean, estimated from replicate PCR estimates for population copy number (n=3) or from means of individual colony estimates (n=15) in the case of colony level estimates. No colony level estimates are shown for sexual population 2 at transfer 0 and asexual population 7 from transfer 20 as no colonies from these populations successfully grew.

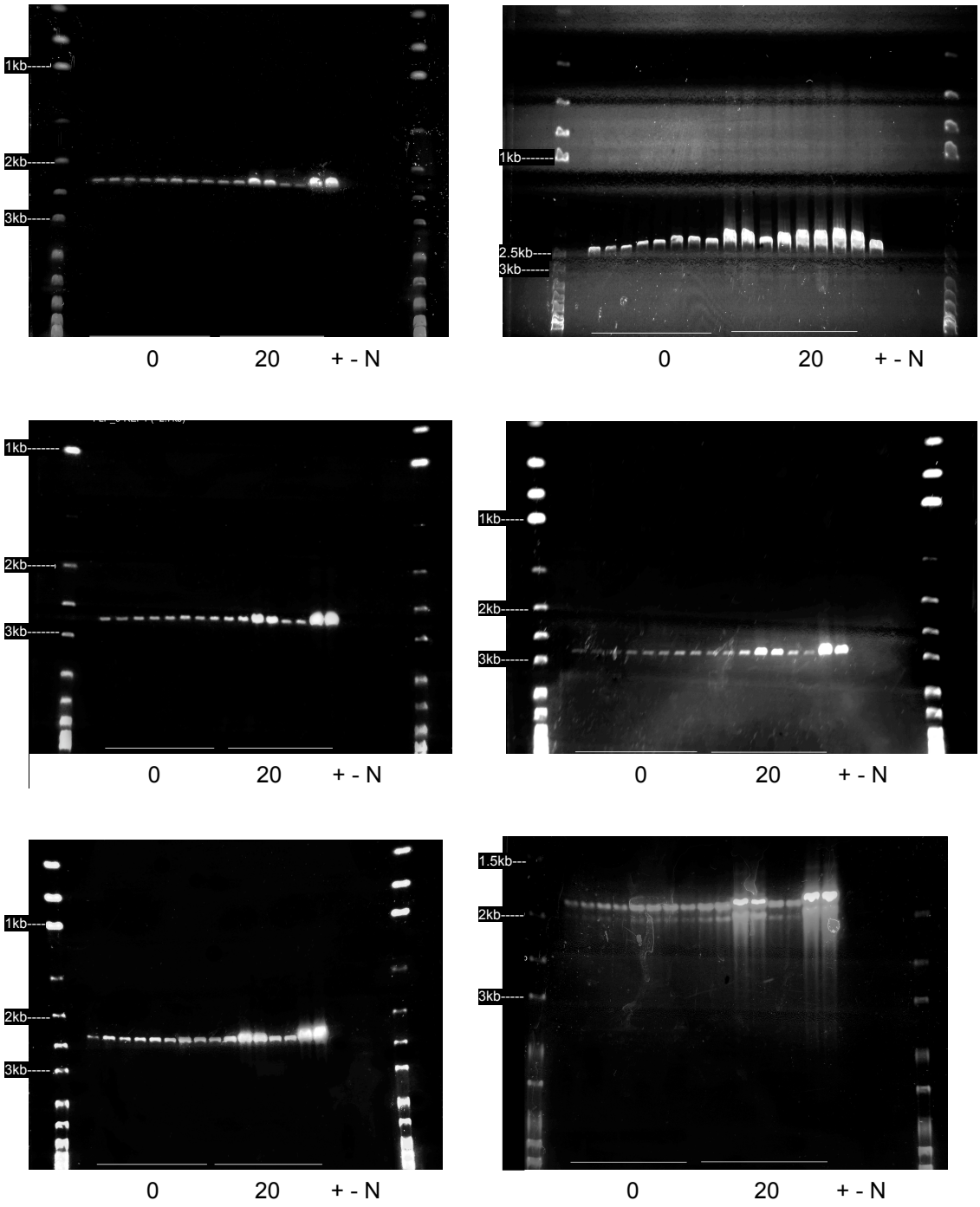


Figure 3.3.3. PCR fragments from sexual populations at transfers 0 and 20 to check for genome rearrangements. A positive control (+), negative control (-) and no template control (N) are also included. Plasmid was amplified between primers (a) STB_2 for – REP2_2 rev, (b) FLP for – RAF rev, (c) FLP_3 for – REP1 rev, (d) RAF for – REP2_2 rev, (e) REP2_2 for – FLP_3 rev, (f) REP1 for – STB_2 rev. PCR fragments are shown along side Hyper ladder III (Bioline).

DNA extractions from the population level estimates were also checked for changes in genome size by multiple PCR reactions. No noticeable changes in fragment size were observed, therefore there is no evidence of genome rearrangements, insertions or deletions (figure 3.3.3). It should be noted however that, as PCR was performed on population level DNA extractions, genotypes at low frequency in the population may be missed.

3.3.2 The invasion of mitochondrial mutants

Respiratory deficient mutants were recorded in all populations within each treatment. At transfer 0 between 0 – 7% of colonies were respiratory deficient with no difference between treatments (ANOVA Treatment: $F_{2,18}$=1.503, p=0.249, N=24, table 3.3.7). Over the course of the experiment respiratory mutants increased dramatically in the sexual treatment, but remained low in the asexual treatment (Figure 3.3.4; repeated measures MANOVA Transfer x Treatment: $F_{10,36}$=5.921, p<0.001, N=24, table 3.3.6). In both the Asexual and Budding control treatments mutant frequency remained below the initial maximum of 7% in all populations throughout the 22 transfers and no effect of transfer

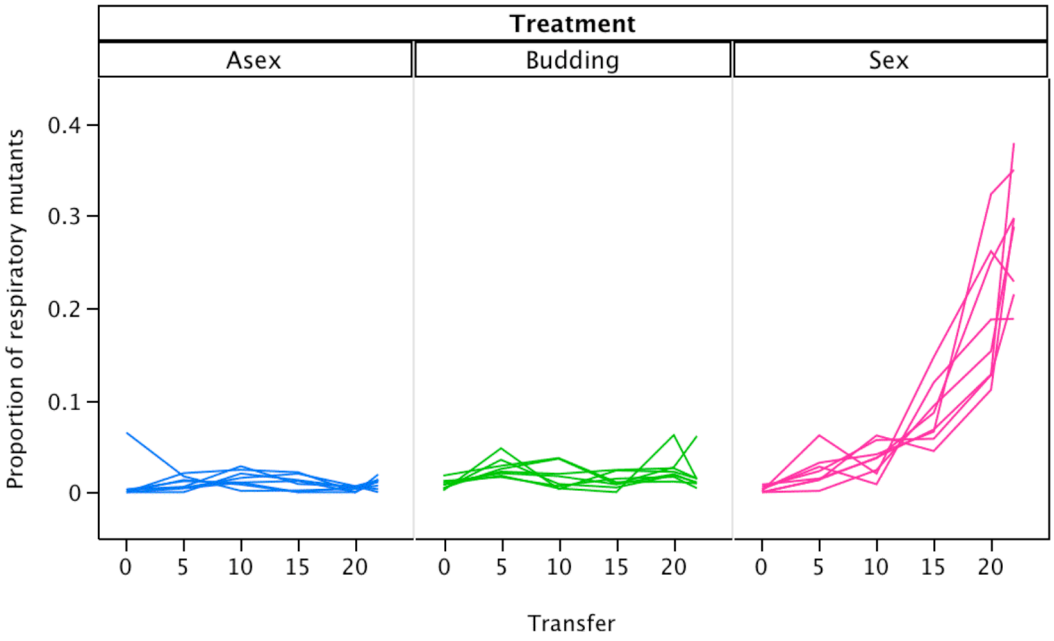


Figure 3.3.4 Proportion of respiratory deficient mutants in over time in reproductive treatments, sex (pink), budding (green) and asexual (blue). Proportions are based on the number of colonies unable to grow on glycerol, a non-fermentable carbon source, among between 500 and 1000 colonies.

Source	Num df	Den df	*F*	*p*
Between subjects				
Treatment	2	21	163.477	<.0001*
Within Subjects				
Transfer	5	17	28.828	<.0001*
Transfer* Treatment	10	36	5.921	<.0001*
N=24				

Table 3.3.6 Repeated measures MANOVA on the proportion of respiratory mutants in each population (arcsine root transformed). Colonies were scored for ability to grow on glycerol media at transfers 0, 5, 10, 15, 20 and 22.

Transfer	Source	df	Mean sq	*F*	*p*
0	Treatment	2	0.00485	1.503	0.2492
	Population[Treatment]	3	0.00314	0.975	0.4265
	Error	18	0.00322		
			***F*=1.186**	**p=0.355**	**N=24**
5	Treatment	2	0.0113	6.683	0.0068*
	Population[Treatment]	3	0.00562	3.334	0.0428
	Error	18	0.00169		
			***F*=4.673**	**p=0.0066**	**N=24**
10	Treatment	2	0.0115	4.315	0.0295*
	Population[Treatment]	3	0.000790	0.296	0.8279
	Error	18	0.00267		
			***F*=1.903**	**p=0.1438**	**N=24**
15	Treatment	2	0.104	32.472	<.0001*
	Population[Treatment]	3	0.00228	0.710	0.5584
	Error	18	0.00321		
			***F*=13.415**	**p<.0001**	**N=24**
20	Treatment	2	0.319	83.232	<.0001*
	Population[Treatment]	3	0.00279	0.727	0.5491
	Error	18	0.00384		
			***F*=33.729**	**p<.0001**	**N=24**
22	Treatment	2	0.4806	242.015	<.0001*
	Population[Treatment]	3	0.00932	4.693	0.0137
	Error	18	0.00199		
			***F*=99.62**	**p<.0001**	**N=24**

Table 3.3.7. ANOVAs for the frequency of respiratory mutants (arcsine root transformed) at each transfer measured. Using the Bonferroni correction, significance is accepted below p=0.0083.

Source	Num df	Den df	*F*	*p*
Sexual	5	3	43.654	0.0053*
Budding	5	3	9.124	0.0492*
Asexual	5	3	5.262	0.1011
N=8				

Table 3.3.8. Multiple repeated measures MANOVAs for the effect of transfer on proportion of respiratory mutants (arcsine root transformed) within each treatment. Alpha level adjusted to p=0.0167 according to Bonferroni correction.

was found when each treatment was analysed independently (repeated measures MANOVA Transfer (budding): $F_{5,3}$=9.124, p=0.0492, N=8 and (asexual) $F_{5,3}$=5.262, p=0.1011, N=8 with alpha level adjusted to p=0.0167, table 3.3.8). In the Sexual treatment mutant frequency increased in all populations during the experiment (repeated measures MANOVA Transfer: $F_{5,3}$=43.654, p=0.0053, N=8, table 3.3.8). Up to transfer 10, respiratory mutants remained at similar frequencies to the control treatments (ANOVA Treatment: $F_{2,18}$=4.315, p=0.295, N=24; table 3.3.7). From transfer 15 onwards the percentage of mutants clearly increases, as shown in figure 3.3.4, and by the end of the experiment 18 – 38% of colonies in the sexual populations were unable to respire.

The impact of sexual reproduction on the cellular ecology of these populations is illustrated by figure 3.3.5. While all treatments contained similar amounts of both selfish genetic elements at the start of the experiment, sexual populations showed both an increase in copy number per cell of the 2µm plasmid, and in the prevalence of selfish mitochondria after just 20 sexual generations. There was no discernible correlation between the plasmid copy number and mitochondrial frequency (Pearson product-moment correlation: r=-0.0549, p=0.7989, N=24), when residuals of a one-way ANOVA with treatment were compared (copy number square root transformed, respiratory mutants arsine root transformed for normality). This lack of relationship was confirmed by comparing colonies with and without these mitochondrial mutants for differences in copy number. Among the six evolved populations which contained respiratory mutations

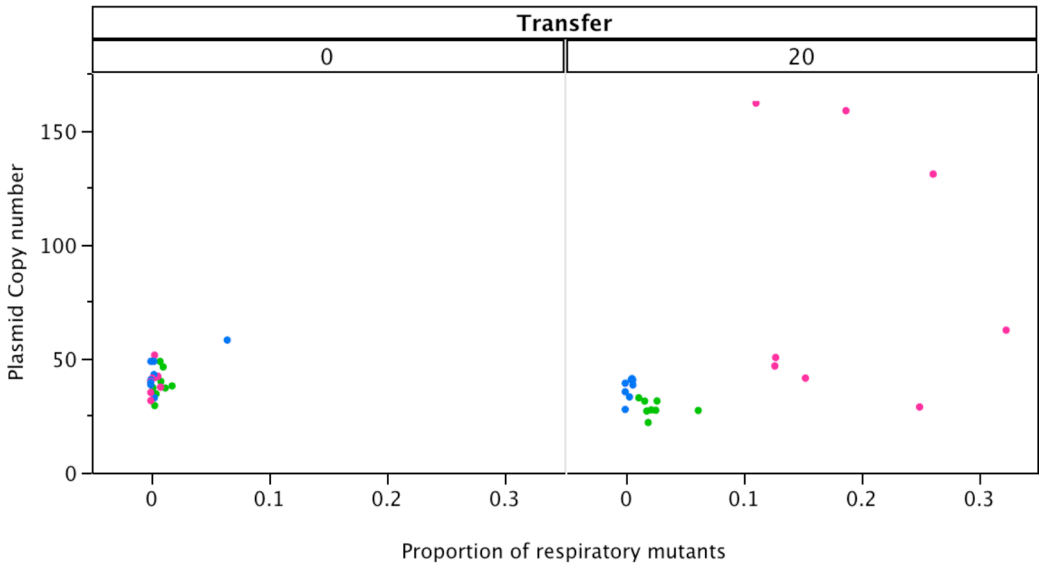


Figure 3.3.5 Over 20 transfers sexual populations (pink) increased in copy number of the 2µm plasmid and/or increased in the proportion of mitochondrial mutants, whilst asexual (blue) and budding (green) populations remained constant.

Source	df	Mean Sq	*F*	*p*
Transfer	5	0.237	1.146	0.346
Respiratory Function	1	0.0202	0.0979	0.756
Transfer*Respiratory Function	5	0.0546	0.264	0.931
Error	60	0.206		
			***F*=4.377, *p*=209, *N*=72**	

Table 3.3.9 ANOVA for the effect of respiratory function on copy number (log transformed). Data are derived from 15 individual colonies in each of the 8 replicate populations at transfers 0 and 20 (excluding colonies which failed to grow).

at transfer 20, no effect of respiratory function on copy number (normalised using log10 transformation) was detectible (ANOVA respiratory function: $F_{1,198}$=0.202, p=0.264, N=72, table 3.3.9).

3.3.3 Fitness

Growth rate and saturation density was measured in 15 colonies from all populations at transfers 0 and 20, as mentioned above. Measuring growth rate in colonies, rather than whole populations, has the advantage of avoiding competition between genotypes occurring during the growth assay.

Growth rate (maxV) and saturation density (maxOD) measurements were strongly correlated (Pearson product moment correlation: r=0.803, p<.0001, N=660; table 3.3.10, figure 3.3.6). Residuals obtained using an ANOVA with treatment, population[treatment] and transfer as main effects were plotted to remove confounding effect of these factors. A slope was calculated using Reduced Major Axis (RMA) regression (Sokal & Rohlf 1981) which found the relationship between maxOD and maxV to be close to 1 (RMA regression: Slope=1.102, 95% CLs=1.041 - 1.166, N=661; table 3.3.10). Therefore only growth rate was used for further analysis.

In transfer 0, population growth rate was the same for Sexual and Budding treatment, but significantly lower in the asexual treatment (ANOVA treatment: $F_{2,314}$=14.943, p<.0001, N=339; Tukey HSD showed asexual treatment to be different to sexual and budding treatments, table 3.3.11). This indicates a difference between strains S and A at the start of the experiment. There was also a significant effect of population (nested within treatment) (ANOVA population[treatment]: $F_{2,314}$=3.976, p<.0001, N=339 table 3.3.11).

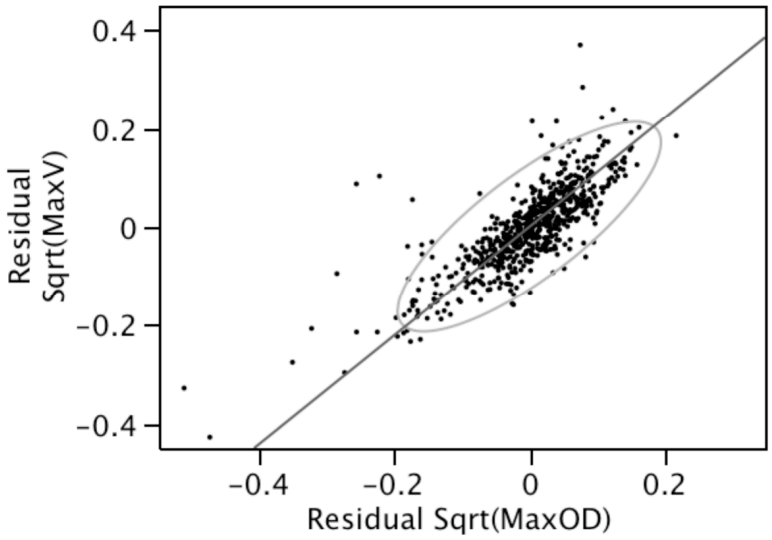


Figure 3.3.6 Reduced major axis regression for the pairwise correlations between growth rate (maxV) and saturation density (maxOD). Variation due to treatment, population[treatment] and transfer was removed using ANOVA with these factors as main effects and comparing residuals. Covariance was analysed using Pearson product-moment correlation and the relationship defined using an RMA regression. 95% density ellipses shown in grey.

X	Y	Variance Ratio	Intercept	Slope	Lower 95%	Upper 95%	r	*p*
Residual Sqrt MaxOD	Residual Sqrt MaxV	1.090	-7.1e-16	1.102	1.041	1.166	0.803	<.0001*

N=661

Table 3.3.10 Test of correlation between MaxV and MaxOD. Data were square root transformed and fitted to an ANOVA with treatment, populations[treatment], transfer and respiratory function as factors. Residuals from this analysis were used for comparison. Regression line was calculated using a reduced major axis regression, using ratio between standard deviations to estimate variance ratio. Pearson product-moment correlation was used test significance of correlation.

Source	df	Mean Sq	*F*	*p*
Treatment	2	0.0874	14.222	<.0001*
Population[Treatment]	21	0.0233	3.790	<.0001*
Error	315	0.00615		

***F*=9.6597, *p*<.0001, *N*=339**

Figure 3.3.11 ANOVA conducted on growth rate (maxV) in populations at transfer 0. Tukey HSD test found the asexual treatment to be different to the sexual and budding treatments in both analyses.

Relative fitness (w) was calculated as the growth rate of evolved populations at transfer 20, compared to mean ancestral growth rates. Therefore a fitness of 1 represents no change from the ancestor, while >1 is an increase in fitness. Fitness was assayed in 15 genotypes within each population, and the mean for each population calculated from the average of these 15 genotypes. All populations showed an increase in fitness over the course of the experiment, with the exception of 2 budding populations, which decreased slightly relative to the ancestor (w=0.924 and 0.980). One sexual population was excluded from this analysis as 14 out of 15 ancestral colonies failed to grow, and the remaining colony was respiratory deficient and therefore unlikely to be a good estimate of population growth rate at the start of the experiment. Due to these exceptional circumstances, this population was removed from further analysis as an outlier. Across all treatments fitness increased by 67% (mean w=1.672, s.e=0.0462, N=345), relative to the ancestral populations. An ANOVA was conducted on log(fitness) (transformed for normality) and found main effects of both treatment (ANOVA treatment: $F_{2,20}$=3.741, p=0.0249, N=307) and population (ANOVA population[treatment]: $F_{2,20}$=11.663, p<.0001, N=307; table 3.3.12). The difference in treatments was attributable to a significant difference between the asexual and budding treatments (figure 3.3.7), where asexual populations showed a much larger increase in relative fitness (mean w(asexual)=1.950, s.e.= 0.0838, N=104, and mean w(budding)=1.277, s.e.= 0.0588, N=120). The mean fitness of sexual populations was most similar to asexual treatments (mean w(asexual)=1.893, s.e.=0.0816, N=83) but was not significantly different from either asexual nor budding.

No correlations were identified between fitness, plasmid copy number and the frequency of selfish mitochondria among the evolved sexual populations when analysed at the populations level (table 3.3.13). The effect of these two selfish genetic elements was also addressed at the within population level. Data for copy number and fitness (square root transformed for normality) were analysed using an ANOVA with treatment, population[treatment], plate and respiratory function as main effects and the residuals plotted (figure 3.3.8). Copy number (square root transformed) showed a weak but significant correlation with fitness (square root) (Pearson product-moment correlation: r= -0.12, p=0.0375, N=345) with a shallow negative slope (RMA regression: Slope=-0.02694, 95% CLs= -0.054 – -0.0001; table 3.3.14). Therefore, this slope represents a ~0.07% (CIs= 0.29 – 0.00001) decline in fitness per additional plasmid.

No appreciable effect of mitochondrial mutants was observed, either between or within populations. Of the 339 colonies assayed from ancestral populations, only four showed respiratory deficiency (indicating a loss of mitochondrial function). In these colonies growth rate was lower on average when compared to respiratory competent colonies

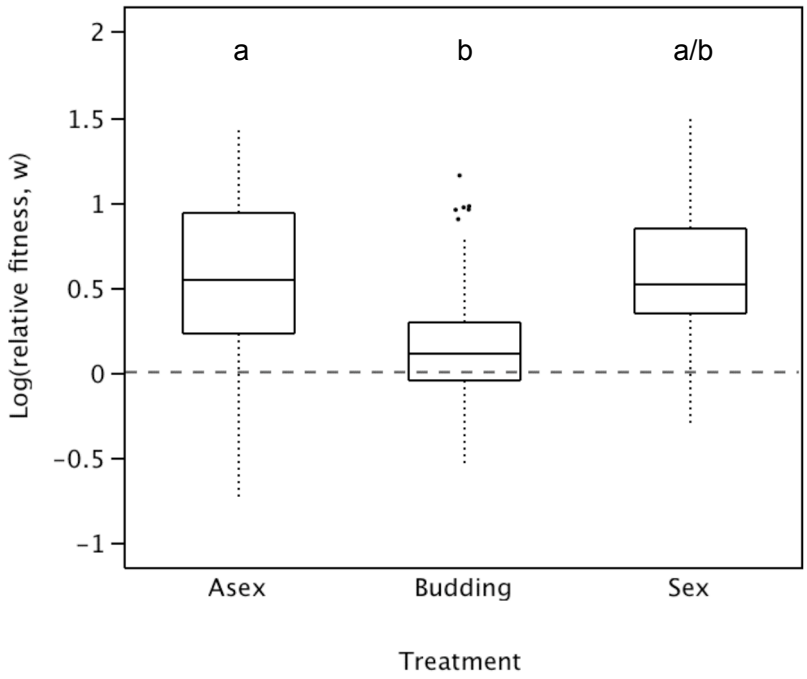


Figure 3.3.7 Effect of reproductive strategy on fitness of evolved populations, Fitness was calculated as the growth rate of evolved cultures, relative to ancestral growth rates. The natural logarithm of fitness is presented, therefore 0 represents no change from the ancestor. Significance between treatments is denoted by lettering.

Source	df	Mean Sq	*F*	*p*
Treatment	2	0.346	3.741	0.0249*
Population[Treatment]	19	1.080	11.663	<.0001*
Error	285	0.0926		

***F*=16.512, *p*<.0001*, *N*=307**

Table 3.3.12 ANOVA for the effect of treatment and populations (nested within treatment) on fitness following the removal of one sexual population identified as an outlier. Fitness was log transformed prior to analysis to normalise resisduals. Tukeys HSD test found the asexual and budding treatments to be statistically different (test slices between asexual and budding treatments: $F_{1,285}$=7.481, p=0.007, N=307). The sexual treatment showed no significant difference to either treatments.

X	Y	Lower 95%	Upper 95%	r	*p*
arsine(Selfish mitochondrial freq)	Fitness	-0.599	0.786	0.183	0.665
Plasmid copy no.	Fitness	-0.604	0.783	0.175	0.679
Plasmid copy no.	arcsine(Selfish mitochondrial freq)	-0.755	0.647	-0.1076	0.801

N=8

Table 3.3.13 Correlations between plasmid copy number, frequency of selfish mitochondria and fitness within evolved sexual populations.

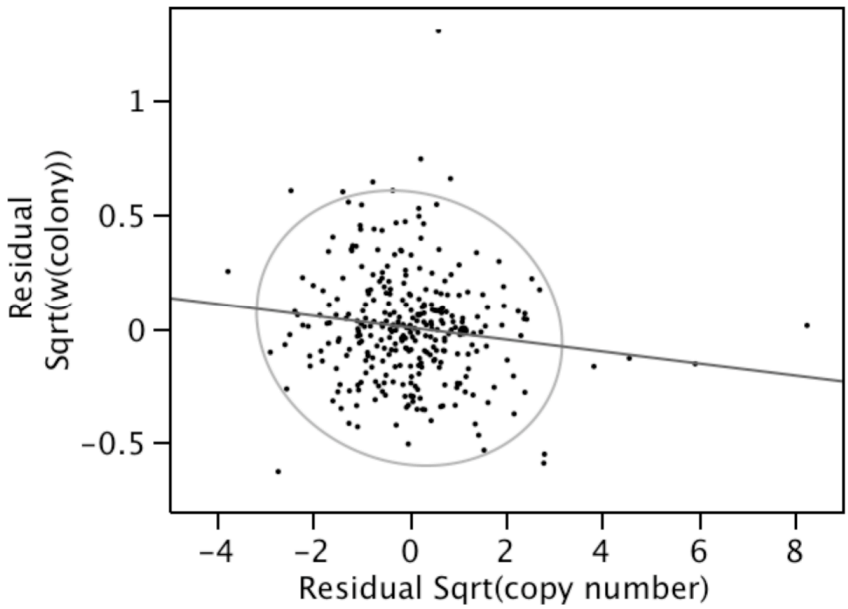


Figure 3.3.8 Reduced major axis regression for the pairwise correlations between growth rate (maxV) and copy number. Variation due to treatment, population[treatment] and respiratory functions was removed using ANOVA with these factors as main effects and comparing residuals. Covariance was analysed using Pearson product-moment correlation and the relationship defined using an RMA regression. 95% density ellipses shown in grey.

X	Y	Variance Ratio	Intercept	Slope	Lower 95%	Upper 95%	r	*p*
Residual Sqrt copy number	Residual Sqrt fitness	0.195	1.96e-16	-0.02694	-0.054	-0.0001	-0.12	0.0375

N=345

Table 3.3.14 Description of regression to test for the effect of copy number on fitness in evolved populations. Data were square root transformed and fitted to an ANOVA with treatment, populations[treatment], plate and respiratory function as factors. Residuals from this analysis were used for comparison. Regression line was calculated using a reduced major axis regression, using ratio between standard deviations to estimate variance ratio. Pearson product-moment correlation was used test significance of correlation

(maxV(respiratory deficient)=0.353, s.e.=0.0951, N=4 and maxV(respiratory competent)=0.475, s.e.=0.00713, N=335). In evolved populations however the reverse was true (maxV(respiratory deficient)=0.798, s.e.=0.0381, N=20 and maxV(respiratory competent)=0.632, s.e.=0.00925, N=302), although respiratory deficient growth rates were within the range observed for competent colonies. In terms of fitness difference,

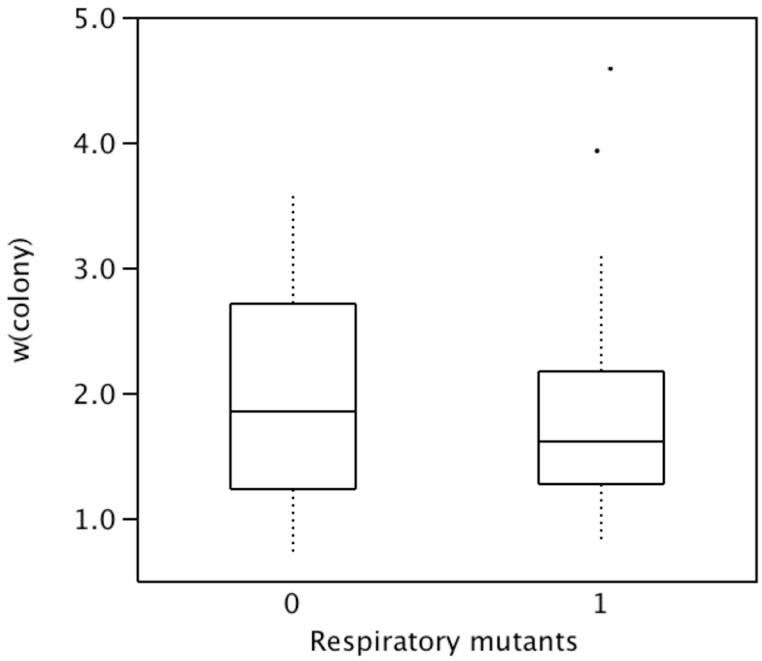


Figure 3.3.9 Fitness was estimated for the respiratory deficient (0) and competent (1) subsets of each population in the sexual treatment. Fitness was calculated as the mean growth rate of each group per population relative to the mean growth rate for each population at the start of the experiment.

Source	df	Mean Sq	*F*	*p*
Population	4	2.598	11.115	<.0001*
Respiratory mutants	1	0.368	1.576	0.216
Population*Respiratory mutants	4	0.309	1.321	0.276
Error	198	0.0637		
			***F*=11.827, *p*<.0001*, *N*=57**	

Table 3.3.15 ANOVA for the effect of respiratory function on within population fitness. Relative fitness of colonies within populations containing both respiratory mutants and respiratory competent colonies were analysed.

colony level estimates were analysed from the populations which contained both respiratory mutants and respiratory competent colonies (five sexual treatment populations; figure 3.3.9). There was a significant main effect of population (ANOVA population: $F_{1,47}$=11.115, p<.0001, N=57, table 3.3.15), but not of respiratory function (ANOVA respiratory function: $F_{1,47}$=1.576, p=0.216, N=57).

3.4 Discussion

3.4.1 Sex supports the evolution of selfish genetics elements

Sexual reproduction, and specifically outcrossing, supports the evolution of selfish genetic elements, by providing an environment under which intragenomic conflicts can arise. This is achieved in two ways. Firstly, by breaking up the genome through recombination, making genes, or linked groups of genes, units of selection themselves (Hurst 1992). Secondly, sex allows costly elements able to drive (transmit to >50% of offspring) to spread through the population, imposing a cost on the individual (Futcher et al. 1988). This has been previously demonstrated experimentally in three different selfish genetic elements of *S. cerevisiae,* where elements are able to invade sexual, but not asexual or inbred populations (Futcher & Cox 1983; Goddard et al. 2001; Zeyl et al. 1996). Therefore it follows that the amount of sex will determine the extent of the cost elements are able to impose (Bestor 1999). Our experiment has demonstrated the key role sex plays in the evolution of two such elements, which between them illustrate the ability of sex to drive both the creation, and increasing virulence, of selfish genetic elements.

2 micron plasmid copy number

Sexual populations showed a substantial change in copy number over the course of the experiment. In the most extreme population, copy number increased from ~40 copies per cell, to ~190, which represents a 2.8% increase in the total cellular DNA. Changes in copy number were not uniform however, and varied both through the course of the experiment and between populations. During the initial 10 transfers all populations showed a modest increase in copy number. In the subsequent transfers, between 10 and 20, the populations differed significantly, with three increasing to three times their original copy number while others remained more stable or declined, suggesting that, in some populations, high copy number plasmids have emerged, while in others antagonistic coevolution between the plasmid and cellular suppression mechanisms have evolved.

In contrast to the sexual populations, under both asexual treatments copy numbers either declined or remained constant. The budding treatment represents the most similar situation to that under which the yeast/plasmid relationship has evolved. In natural and laboratory strains of yeast outcrossing occurs approximately once every 50,000 generations (Ruderfer et al. 2006), therefore it is unsurprising that no significant change in copy number was observed over the ~150 generations of this experiment. In the asexual treatment copy number did change significantly over the 20 transfers. Unlike the budding treatment, several of the asexual populations increased slightly in copy number following an initial decline. The asexual treatment differed from the latter in that it was put through the process of sporulation, to produce clonal diploid spores. This difference in dynamic between the two asexual treatment may be due to the mutagenic nature of the sporulation process, as starvation is known to increase mutation rate in *S. cerevisiae* (Marini et al. 1999), increasing the pace of adaptation for both plasmid and genome (Springman et al. 2009). The inclusion of this treatment demonstrates that changes observed in the sexual treatment are not due to this potentially confounding effect of increased mutation rate. Despite a significant change in copy number over the 20 transfers, copy number did not significantly rise in any asexual population.

Two sources of selection may be responsible for driving high copy number. Firstly, there may be a direct benefit for the plasmid in having increasingly high copy number. The mostly likely source of selection would be to maximise plasmid stability, i.e. the fidelity of plasmid transfer into the next generation. Plasmid free ([cir^0]) cells have between a 1.5 – 3% growth rate advantage over those containing the plasmid ([cir^+]) (Mead et al. 1986). Therefore [cir^0] cells can increase in the population both through spontaneous loss of the plasmid and through selection (Futcher et al. 1988). As stability has been shown to correlate positively with increasing plasmid copies (Futcher & Cox 1984), this may suggest an advantage to having high copy number. However, the plasmid employs a method of hitchhiking on the cells own chromosomal segregation system (Hajra et al. 2006), making the probability of loss through miss-segregation fairly low, ~5.7×10^{-5} per generation in haploids (Futcher & Cox 1983). Futcher et al (1998) estimated that rates of outcrossed sex as low as 2×10^{-4} are sufficient to prevent the spread of [cir^0] cells into the population, compared to our experimental rate of ~0.14, and as we did not observe any loss of the plasmid from any populations, spontaneous loss is unlikely to represent a significant selective pressure

A second source of selection for high copy number may be driven by within-individual competition between elements. As each cell represents a closed environment with finite resources, plasmids able to use up a higher proportion of those resources will have a selective advantage over rival genotypes. Therefore, we would expect competition to

favour increasingly aggressive replicators. The increase in sexual reproduction and, as a consequence, genetic mixing will increase the number of competing genotypes with in the cell, intensifying competition. Different variants of the plasmid have been shown to coexist and co-segregate within a cell (Livingston 1977), although the dynamics of multiple plasmid carriage in the long term is yet to be investigated.

One surprising result highlighted by this experiment was that estimates of 2μm copy number taken from individual colonies were not necessarily representative of the copy number estimated from whole populations. There may be several alternative explanations for this. The first is that high copy number populations contain super numary plasmids at low frequency. Extremely high copy numbers, up to 750 per diploid, have been observed in other yeast strains (see chapter 4, this thesis), therefore it is not unreasonable that such plasmids may have evolved in these populations and be found at frequencies less than those detectable from the 15 colonies sampled. However to achieve the difference between means seen here, super numary plasmids would need to reach >1500 copies per cell in order to be missed. What's more, 3 populations showed this skew, making the probability of not detecting a super numary plasmid much lower, and subsequently the minimum copy number several orders of magnitude higher, at around 330,000 copies!

It is possible that a sampling bias existed towards lower copy number colonies if a. high copy number cells are more likely to die during the freezing and thawing process and b. low growth rates in these variants made colonies smaller and less likely to be picked. Population level estimates were taken from DNA extracted directly from frozen population stocks, while colony level estimates were plated and picked from agar plates for extraction. As the freeze-thaw process does lead to some cell death due to cold and osmotic-stress (Calcott & Rose 1982), any cells not able grow following the freeze-thaw process would not be selected for colony level extraction, but would be included in the population DNA pool (which did not depend on cell growth). Of the cells able to form colonies, we used a grid system to minimise sampling bias, as colonies nearest each cross on the grid are chosen. However, low growth rates can increase the chance of colonies going extinct early in their growth and therefore being undetected (Martin et al. 2010). Therefore if plasmids at high copy number increase the probability of colonies failing to form through either of these processes, this would bias our estimates at the colony towards low copy number plasmids.

Another possibility is that copy number has altered between these two estimates. While population level copy number estimates represent the plasmid titre at the time of freezing, colony level assays were allowed to form colonies and then had 3-4 rounds of mitotic growth

in culture during growth plate reading. Copy number is known to fluctuate within clones, possibly due to uneven partitioning between dividing cells (Futcher & Cox 1984). The plasmid's behaviour during meiosis, or in haploids vs. diploids, is currently unknown and it is not inconceivable that during, or following, sexual reproduction, aggressively replicating plasmid variants are able to reach higher copy numbers, but are more successfully repressed during mitotic growth or in diploids. If this was to be the case, such an adaptation would be advantageous to the plasmid, as the fitness burden on the individual of being at high copy number would only be borne under circumstances in which it provides a transmission advantage.

Selfish mitochondria

In addition to the changes in 2μm copy number, we also chose to screen our populations for the presence of selfish mitochondria. Mitochondrial genomes are known to have relatively high mutation rates, including partial deletions of the genome (Dujon 1981; Zeyl & DeVisser 2001). Some such mutations result in the complete loss of respiratory function. As yeast is also capable of fermentation, this is not fatal, however it does result in ~30% selective disadvantage relative to respiratory competent cells (Zeyl & DeVisser 2001). A common class of mitochondrial mutants involve the deletion of roughly half of the genome and the duplication of the remaining portion (Dujon 1981). Where this involves the origin of replication (*ori*) this gives the symbiont a within-individual replication advantage. These mutants are referred to as suppressive or hyper-suppressive depending on their ability to out replicate other mitochondria. In large asexual populations selection on the individual quickly removes these mutants from the population (Taylor et al. 2002). However, as yeast are isogamous and have biparental inheritance of mitochondria, these 'selfish' mitochondria are effectively able to drive and so, like the 2μm, should be able to spread in sexually reproducing populations. These data clearly show the ability of sex to facilitate the spread of such elements, as all populations in the sexual treatment reach high (between 20-40%) frequencies of respiratory incompetent cells, while all asexual and budding populations maintained a steady frequency of ~1%. Therefore through just 22 sexual generations out of a total of ~150, sex has turned one of the most fundamental eukaryotic endosymbionts, into a parasitic element on its way to fixation in the population.

3.4.2 Sex did not alter the scale of adaptation in evolved populations

The fitness of evolved populations was calculated as the growth rate of the populations at the end of the experiment, relative to their ancestor. Saturation density was also measured but showed a close correlation to growth rate. Therefore both would provide a similar estimate of fitness in these populations. Fitness increased in all but two populations, with

very large variation between populations. This increase was much greater in asexual and sexual treatments, which showed an 85% and 73% increase in fitness respectively compared to just 21% in the budding treatment. However, only the difference between the asexual and budding treatments was significant. As mentioned previously, sporulation increases mutation rate due to the effects of starvation (Marini et al. 1999), and provides the most likely explanation for the difference in adaptation between asexual and budding populations. Therefore, it is probable that the asexual (and also the sexual) treatment has experienced a higher rate of adaptation than the budding treatment (Springman et al. 2009). Our finding that sexual populations did not differ from that of the asexuals, and in fact appeared slightly lower, is in contrast to the results of Goddard et al. (2005). In this experiment the authors evolved similar *S. cerevisiae* populations under sexual and asexual conditions, and in harsh and benign environments. Following 100 generations under 'harsh' (low nutrient, high temperature) conditions, asexual fitness increased by 30%, and sexual, 40%, a significant difference of 10% between treatments, while benign populations showed no adaptation to their environment. The difference in the scale is likely to be environment and/or protocol specific, however while sex appeared to be beneficial in the latter experiment, our study found no difference between sexual and asexual populations.

As the frequency of sex was very different between treatments (every 25 generations in Goddard et al (2005) and every 5-6 in this study) it is possible that the adaptive benefits of sex are being negated due to the increasing burden of intragenomic conflict driven by the increased regularity of sexual reproduction. This increasing conflict may cause sex to be costly in two ways; through the direct burden of increasingly virulent elements, and also through the cost of coevolution to counter increasing element virulence.

3.4.3 The extent of selfish genetic element burden does not predict the fitness of evolved populations

Between populations, fitness variation could not be explained by plasmid, or selfish mitochondrial burden. However, within populations we observed a fitness cost of ~0.07% per plasmid, which similar to the 0.09% fitness cost estimated in Chapter 2. However confidence limits around this estimate were extremely wide. This large variability is not surprising. Adaptations in the host may emerge to either suppress or tolerate increasing genetic conflict in the cell. Adaptations for resistance to selfish genetic elements may themselves be costly (Boots & Begon 1993; Kraaijeveld et al. 2002), therefore increasing the cost per element, while alleviating mutations may make the cost of selfish genetic elements relatively cheap. Therefore, under a coevolution model we expect different adaptations for the control or mitigation of intragenomic conflict to appear in different populations, meaning that the cost of selfish genetic elements will have diverged by the end

of the experiment. However within a population, or genotype, an increase in burden is still expected to increase the cost to the host.

Surprisingly, no effect of selfish mitochondria on fitness was detected, despite the expected fitness cost of ~30%. There are several possible explanations for this outcome. Firstly it is possible that the assay environment was sufficiently different from the selective environment and insufficient to detect the cost of lacking mitochondria. As growth assays were performed in small 200µl cultures it is unlikely that there would be have been a large amount of mixing during growth, therefore potentially allowing oxygen depletion in the bottom of the growth plate. In the absence of oxygen all cells use fermentation for energy metabolism, which would negate the benefit of having functional mitochondria for respiration. However, mitochondria are also important during fermentation, and the lost of mitochondrial function results in decreasing alcohol tolerance and inhibits growth (Hutter & Oliver 1998). Therefore we would still expect respiratory deficient cells to be at a disadvantage in this environment. Secondly, the fitness costs of selfish mitochondria may simply be obscured by large variation in fitness not associated with selfish mitochondria. Finally, it is also possible that adaptation to compensate for the loss of mitochondrial function has emerged during the course of the experiment. In ancestral populations colonies unable to respire had below average growth rates, compared to colonies with functional mitochondria, while in evolved populations the reverse was true. However, in both cases, mitochondrial mutants fell within the range of growth rates seen in respiratory competent colonies, making it difficult to infer adaptation in the scope of this assay.

3.4.4 Antagonistic coevolution negates the benefits of sex

In summary, we have determined that over the course of approximately 150 generations, with a frequent rate of sexual reproduction (one meiotic to six mitotic generations), sex has greatly increased the intensity of intragenomic conflict when compared to asexually growing counterparts. In contradiction to previous results obtained in the same system, but with a low rate of sex (one meiotic to 25 mitotic generations), we found no adaptive benefit to sex at the end of our experiment. The beneficial nature of sexual reproduction, through breaking down linkage disequilibrium, is in fact dependant on virtually any level of recombination above zero, and there is little benefit to increasing recombination in this respect (Charlesworth et al. 1993; Haddrill et al. 2007). Therefore, as the rate of sex increases, the growing intragenomic conflict may be sufficient to reduce, and potentially even negate the benefits of sexual reproduction. This would be achieved in two ways. First and most simply, we have shown that the burden of selfish genetic elements increases in sexual populations. If we compare the expected costs of the 2µm and selfish mitochondria on populations at the end of the experiment, sexual populations would experience an average burden of 3.6% for

carriage of additional plasmids (Chapter 4, this thesis) and 5.7% from the increasing the frequency of selfish mitochondria in the population (Zeyl & DeVisser 2001). In asexual populations both 2μm copy number and selfish mitochondria decreased slightly. Therefore by this basic calculation, fitness should be increased by 0.9% relative to ancestral populations. In the experiment by Goddard et al (2005) sexual populations showed a relative increase in fitness around 10% greater than sexual populations, on a similar scale to the predicted 9.3% cost of these two selfish genetic elements.

However, the costs of carrying selfish genetic elements are unlikely to be fixed, as adaptations for tolerance or resistance will lead to rapid divergence in the relative cost of having these elements. Coevolution of bacterial-plasmid associations has repeatedly demonstrated that over time, either the host (Bouma & Lenski 1988; Modi & Adams 1991), or the host and plasmid (Dahlberg & Chao 2003; Dionisio et al. 2005) evolve to reduce the cost of plasmid carriage. Therefore, we can expect the individual to evolve compensatory or resistance adaptations to increasingly costly elements, however these adaptations may themselves impose a fitness cost (Lenski 1988a; Lenski 1988b). The cost of resistance will also vary greatly depending on the specific mutation (Lenski 1988a), and trade-offs can themselves be ameliorated (Lenski 1988b). The fitness trajectories can therefore be extremely complex and it is unsurprising that, in this study, neither plasmid copy number, nor selfish genetic element frequency, predicted the difference in fitness of populations. In our sexual populations we observed two broadly different responses of plasmid copy number evolution; a >3x large increase in copy number, and an initial increase followed by copy number reduction. These two outcomes may represent a difference in host response to increasing parasite virulence, to either mitigate the cost of rising copy number, making the maintenance of each plasmid cheaper, or to reduce the plasmid copy number, which may make each plasmid more expensive.

Unfortunately the scope of this study does not allow us to further infer the nature of adaptations in evolved plasmids, mitochondria and host cells. A further extension to this work would be to investigate the extent of coevolution that has occurred in these populations. Three approaches may be taken to improve our understanding of the changes occurring in these populations.

- Fitness estimates of genotypes in the absence of the 2μm and selfish mitochondria, to avoid the confounding fitness effects from other sources. This would provide an estimate of the relative cost of containing the element.
- Measuring fitness of evolved selfish genetic elements in ancestral host populations and vice versa, to measure the extent to which hosts and elements have adapted over the course of the experiment (Buckling & Rainey 2002).

- Sequencing of evolved populations. At present we are only able to broadly infer adaptation in the plasmid, mitochondria and host, however sequence analysis would allow us to not only identify the which of these have evolved, but also the source of this adaptation and the strength and speed of selection. As an example, the posttranslational modification pathways of yeast are known to control 2μm copy number (Chen et al. 2005; Dobson et al. 2005; Sleep et al. 2001; Xu et al. 1999) and would therefore we may expect adaptation in this region to be a source of resistance to plasmid DNA.

3.5 Conclusions

This experiment has illustrated, for the first time, the profound effect of sexual reproduction on the evolution of two different selfish genetic elements and the consequences that this has for the evolutionary benefits of sex. On the one hand we have demonstrated that sex supports the invasion and spread of selfish mitochondria, effectively turning the most fundamental endosymbiont in the eukaryotic cell, into a prolific genetic parasite. In the 2μm, copy number increases were seen in all sexual populations, while all asexually reproducing populations were able to maintain low levels of this element. It should also be considered that this is likely to apply to the many other selfish genetic elements which inhabit the *S. cerevisiae* genome, including transposons, group II introns, RNA viruses, homing endonucleases (Burt & Trivers 2006) and those yet to be identified. In unison, this small army of elements have the potential to become an extremely costly burden to sexually reproducing organisms. In contrast with previous results from this system (Goddard et al. 2005), we found no benefit to sex in terms of adaptation to the environment which may correspond to an increase in the rate of sexual generations in our experiment relative to the latter.

The outcome of this experiment raises the question, not of whether to have sex or not, but how much to have. The benefits of sex are in fact dependant on relatively low rate of recombination (Charlesworth et al. 1993; Haddrill et al. 2007). Here we suggest that the frequency of sexual reproduction determines the intensity of intragenomic conflict, which may reduce its benefits. Indeed, *S. cerevisiae* may be an excellent example of this. As a facultatively sexual species, *S. cerevisiae* reserves sex for times of environmental stress and has evolved several mechanisms to increase the rate of inbreeding, leading to a very low rate of outcrossed sex (Ruderfer et al. 2006). As has been previously proposed (Futcher et al. 1988), the rising burden of selfish genetic elements may be one selective force maintaining this lifestyle.

Despite this, sex has become an extremely widespread mode of reproduction, however this has come at a price. The rapid emergence and spread of selfish mitochondria in our sexual populations demonstrates that the evolution of endosymbionts must have occurred in the absence of sexual reproduction. In evolving to higher rates of sex, mechanisms to quell the rising conflict occurring within the cell, have necessarily evolved. These include several defensive mechanisms to control selfish genetic elements within the genome (Johnson 2007), most universally the evolution of anisogamy may have been driven by intragenomic conflict (Hastings 1992; Hurst & Hamilton 1992). By excluding cytoplasmic elements, such as mitochondria, from one of the gametes intragenomic conflict is prevented, as such elements are essentially being inherited asexually. However, with the emergence of defined sexes results in the twofold cost of sex, therefore suggesting that selection driving the evolution of anisogamy must be strong enough to out-weigh this cost (Hastings 1992).

3.5.1 Future directions

This study has demonstrated experimentally the importance of sexual reproduction in driving intragenomic conflict in a simple, well-mixed environment and demonstrates the strength of using simple systems to test major hypotheses. However the realities of population structure, ecology and behaviour are excluded from this experiment, and provide fertile ground for future investigation. One such avenue would be to investigate the role that sexual selection can play in reducing the intensity of intragenomic conflict.

Despite its apparent simplicity, *S. cerevisiae* is able to exert mate choice using pheromone signal size as an honest signal of quality (Jackson & Hartwell 1990), and this has recently been developed into an excellent model system for studying the genetics of sexual selection (Rogers & Greig 2009). Under non-random mating, the cost of virulence is two-fold, as genetic burden decreases not only the growth rate of the individual, but reduces mating potential, therefore hindering the spread of costly elements through the population. This is analogous to the case made for the effect of sexual selection on sexually transmitted parasites (Knell 1999), but contrary to the predictions for horizontally transmitted parasites (Moller 1996). However, the impact of sexual selection on the evolution of parasitic elements has been largely neglected for the opposing question. This is likely to be due to the ease with which parasitic infections can be manipulated for experimentation. The model system developed by D. Grieg and colleagues provides an excellent system in which populations can be evolved both with and without sexual selection and the experimental testing of these predictions.

Chapter 4

Population Biology of the 2 Micron Plasmid

4.1 Introduction

The 2 Micron plasmid (2μm) of the *Saccharomyces* yeasts is a rare example of a circular DNA plasmid in eukaryotes. Unlike bacterial plasmids, the 2μm has no known functional benefit to the host cell, despite imposing a ~1.5 – 3% reduction in growth rate (Mead et al. 1986). Instead we can consider it a selfish genetic element, which uses the molecular tools of the cell to promote its own replication. The maintenance of this lifestyle is made possible by its transmission to nearly 100% of meiotic progeny when the host reproduces sexually (Futcher et al. 1988). This means that the plasmid can spread through the population. Therefore, this element, and other selfish elements like it, are autonomous units with a population structure and evolution dependent on, but not identical to, their host population. While the molecular biology of the plasmid has been extensively studied (*see for example* Cui et al. 2009; Ghosh et al. 2010; Xiong et al. 2009), very little is known about its population biology.

The 2μm is one of just ten DNA circles found in eukaryotes (Blaisonneau et al. 1997), excluding the rDNA plasmids found in some protozoans and yeasts (Feagin et al. 1992; Huber & Rustchenko 2001). Relative to many bacterial plasmids, the 2μm genome is small (6.4Kb) and compact, containing just four open reading frames (ORFs) and a four *cis* acting regions (Volkert et al. 1989). The structure and function of these sites is has been described Chapter 1 of this thesis. The plasmid exists at between 24 – 88 copies per haploid cell (Gerbaud & Guerineau 1980) and uses a mechanism of rolling circle replication (*see* Chapter 1) in order to maintain high copy number (Futcher 1986), as DNA replication only occurs once per cell cycle, along with regular chromosomal DNA (Zakian et al. 1979). This mechanism is based on recombination occurring at 2 *cis* acting regions on either side of the plasmid, the FLP recognition target (FRT) sites. Site-specific recombination, mediated by the plasmids Flpp protein, between the FRT site on the opposite side of the plasmid allow one side of the plasmid to rotate its orientation relative to the other (*see* Chapter 1, figure 1.2). Therefore, the plasmid consists of two domains, the left unique region (LUR) and right unique region (RUR).

4.1.1 2 Micron Plasmid Variation

The plasmid has been found among nearly all industrial strains of *Saccharomyces cerevisiae* (Gerbaud & Guerineau 1980; Xiao & Rank 1993), and more recently in the wild species, *S. paradoxus* and *S. bayanus* (Nakayashiki et al. 2005). Variation in plasmid copy number and nucleotide sequence has so far only been studied on *S. cerevisiae*, where copy number ranges between 24 and 88 copies per haploid cell (Futcher & Cox 1984; Gerbaud & Guerineau 1980).

Sequence variation has been previously estimated among industrial *S. cerevisiae* strains. To date, two variants of the plasmid have been identified, types 1 and 2, which are defined by ~30% nucleotide divergence at the STB locus (Xiao et al. 1991a), and ~10% divergence in amino acid sequence of the REP1 and REP2 genes on the RUR (Xiao et al. 1991b). Type 1 plasmids were found to be restricted to baking strains, while type 2 plasmids appear to be more variable and found in strains from a number different industrial sources (Xiao & Rank 1993).

4.1.2 Population Biology of Saccharomyces spp.

As a vertically transmitted element, the population structure of the host is extremely important to the plasmid's evolution. The degree of recombination within and between populations ultimately determines the extent to which the plasmid can spread (Futcher et al. 1988), which will influence both its prevalence within populations and its global distribution, and, as we have previously shown, its virulence (*see* Chapter 3). The *Saccharomyces* sensu stricto group has been subject to extensive analysis in the last decade (Liti et al. 2006; Liti et al. 2009). In particular, the wild strain *S. paradoxus* is emerging as a useful model system for microbial population genetics (Johnson et al. 2004; Koufopanou et al. 2006; Tsai et al. 2008). *S. cerevisiae* shows little differentiation on a global scale, most likely as a consequence of its long history of human usage (Liti et al. 2009). This species is a mixture of distinct lineages and 'mosaic' strains containing a patchwork of genomic regions from the 'clean' lineages, and as yet undiscovered lineages (Liti et al. 2009). Strains are often associated with industrial processes; baking, brewing, wine and sake fermentation and distillation, but can also be isolated from clinical sources as well as from the wild.

In contrast, *S. paradoxus* has no known history of use by humans, and strains are isolated universally from oak trees or associated substrates (Koufopanou et al. 2006). The two species share ~85 – 90% nucleotide homology (Kellis et al. 2003; Liti et al. 2006), suggesting a divergence time of between 0.4-3.4 MYA (Liti et al. 2006). Unlike *S. cerevisiae, S. paradoxus* populations show clear spatial structure, with populations from

Europe, the Far East and North America (and one strain from Hawaii) forming clearly defined lineages with no evidence of recombination between them (Koufopanou et al. 2006; Liti et al. 2009). The scale of differentiation between these lineages is much larger than in the globally mixed *S. cerevisiae* population. For instance, the estimated time to most common recent ancestor (TMCRA) for each individual *S. paradoxus* lineage is similar to that of the entire *S. cerevisiae* phylogeny (Liti et al. 2006). The Far Eastern and European populations are the most highly related (~1.4% divergence), with the North American lineage ~4% divergent from either of these (Bensasson et al. 2008).

One aspect of yeast populations of particular importance to plasmid evolution is the rate of outcrossing in the population. We can expect that the rate of outcrossed sex will ultimately determine a. whether or not the plasmid is present in a population, b. how prevalent it is, and c. the virulence of the plasmid. Although highly stable, the plasmid is lost from host cells at a rate of 7.6×10^{-5} from haploid cells (Futcher et al. 1983). As outcrossing allows the plasmid to spread through the population, this will determine the extent to which [cir^0] strains, which have a competitive advantage (Futcher et al. 1988), are able to invade the population, reducing the prevalence of the plasmid containing ([cir^+]) strains, and ultimately causing their loss from the population entirely. As we have demonstrated in chapter 3, outcrossing can also drive the evolution of copy number of the plasmid, which is a measure of the virulence of the element (*see* chapter 2).

4.1.3 Aims

Despite its role as a key yeast vector system and extensive understanding of its behaviour on a molecular scale, precious little is known about the evolution of the 2µm plasmid as a selfish genetic element in its natural population. This chapter aims to build a picture of the population structure and variation of this plasmid by addressing its distribution, variation in copy number and diversity at the sequence level in its host.

4.2 Materials and Methods

4.2.1 Prevalence

Strains of *S. cerevisiae* and *S. paradoxus* were mostly taken from the collection of A. Burt and V. Koufopanou, P. Sniegowski and C.R. Tien. A full list of strains and their sources can be found in table A1, Appendix I.

Strains were grown from glycerol stocks on YPD agar (see appendix I for description) and a single colony inoculated into 1ml YPD for extraction. Following overnight growth

at 30°C cultures were washed in sterile H_2O and the DNA extracted using the Promega Wizard® SV 96 Genomic DNA Purification System as described in appendix I.

The presence of the 2µm was identified using PCR targeting three different plasmid sites, ORFs FLP (FLP_3 for: ACAGCCGACATACATCCCAG, rev: GCGCTTCCGAAAATGCAACG) and REP2 (REP2_2 for: GCTGGTGGGACTAATAACTGTG, rev: ACGACATTGAAACAGCCAAG) and *cis*-acting locus STB (STB_2 for: TGCGTTTCCATCTTGCAC, rev: GAACAGAAATGCAACGCG). All primers were designed from conserved regions of the plasmid identified from sequence data obtained from the Saccharomyces Genome Resequencing Project (SGRP; Carter 2005) and selected to minimize number of SNPs between *S. cerevisiae* and *paradoxus* plasmids. PCR using nuclear primers ITS1 (TCCGTAGGTGAACCTGCGG) and ITS4 (TCCTCCGCTTATTGATATGC) was used as a positive control for DNA quality. PCR was conducted using standard methods described in appendix I.

For further verification of the PCR results, a southern blot was performed using the DIG High Prime DNA Labeling and Detection Starter Kit I from Roche. 20 Strains were screened in total consisting of 5 *S. cerevisiae*, and 5 *S. paradoxus* from the Canadian, European and Far Eastern populations. In addition, type strains YDG598 (*S. cerevisiae*) and CBS432 (*S. paradoxus*) were included along with a no template control. DNA was digested using EcoR1 (New England Biolabs) which produces 2 fragments of 3912 and 2406bp based on the published 2 micron sequence (Genbank. J01347). Blots were performed using a mixed probe for five sites around the 2µm plasmid and a genomic control. Details of the probes, blot and hybridization can be found in appendix I.

4.2.2 Sequencing

Sequences of most *S. cerevisiae* strains (excluding those obtained from C.R. Tien and the *S. paradoxus* strain A12) were obtained from sequence data from the Saccharomyces Genome Resequencing Project (SGRP; Carter 2005). Sequences were collected using BLAST with the published 2µm sequence (strain A634A, Genbank J01347), and then aligned to this in CodonCode Aligner 2.0.6 (CodonCode Corporation, MA, USA). Bases with quality scores <40 were changed to N (Sanger alignments were performed by S. Lomas).

Sequences from *S. paradoxus* strains not included in the SGRP were produced for this study using PCR based sequencing. Sequencing was performed on all European *S. paradoxus,* and North American strains not identified as being clonal to strain A12, plus one clonal strain. The plasmid was amplified in three sections (shown in figure 4.2.1). DNA was amplified using the regular PCR as described in appendix I using an annealing

temperature of 55°C and an extension time of 4 minutes. Fragments were cleaned using Promega Wizard SV Gel/PCR Clean Up and sequenced by Macrogen Korea. Sequencing was initiated from both terminal and internal primers (table 4.2.1).

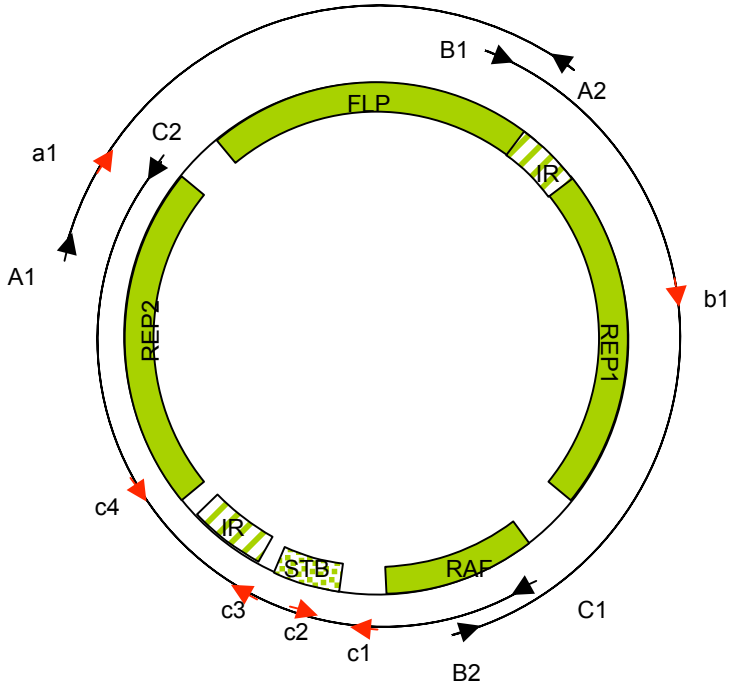


Figure 4.2.1 PCR amplification of the 2µm plasmid for sequencing. Three overlapping sections of the plasmid were amplified using terminal primers (black). PCR products were then cleaned and sequenced using both the terminal primers and additional internal primers (red).

Code	Name	Sequence
A1	REP2_2 for	GCTGGTGGGACTAATAACTGTG
a1	FLP_3 for	ACAGCCGACATACATCCCAG
A2	FLP_3 rev	GCGCTTCCGAAAATGCAACG
B1	FLP for	TGCTTCCTTCAGCACTACC
b1	REP1 for	TCACCCCACAATCCTTCATC
B2	RAF rev	TGCCTTTTGCATTTCTAGTCTC
C1	RAF for	ACGATACACTTCCGCTCAG
c1	STB_2 for	TGCGTTTCCATCTTGCAC
c2	STB_2 rev	GAACAGAAATGCAACGCG
c3	STB for	AATCTTGCACGTCGCATCCC
c4	STB rev	CTTTCGCGTTGCATTTCTGTTC
C2	REP2_2 rev	ACGACATTGAAACAGCCAAG

Table 4.2.1 Sequences of primers used to amplify (bold) and sequence (all) the 2µm plasmid

Contigs were assembled in CodonCode Aligner 2.0.6 by alignment to the published 2µm sequence (a *S. cerevisiae* derived plasmid) or to the *S. paradoxus* A12 plasmid previously aligned from SGRP data.

Analysis

Phylogenies were constructed in Paup 4, using heuristic maximum parsimony searches with a limit of 100 trees. As no outgroup was available, trees were mid-branch rooted unless otherwise stated.

The number of synonymous (S) and nonsynonymous (NS) polymorphic sites where counted manually, with any string of polymorphisms >3bp counted as one. Gaps were ignored. Divergence and polymorphism were calculated manually as the number of fixed differences, or mutations, per site per sequence respectively. dS and dN were calculated as the number of synonymous substitutions per synonymous site per sequence, and nonsynonymous substitutions per nonsynonymous site per sequence respectively. Divergence, polymorphism, dN and dS are multiplied by 100 for clarity.

Macdonald-Kreitman analysis (McDonald & Kreitman 1991) was performed using comparisons of fixed and polymorphic S and NS in a Chi squared 2x2 contigency table. Codon usage bias was calculated using the Codon Adapation Index (Sharp & Li 1987). Calculations were performed in CAIcal, http://genomes.urv.es/CAIcal (Puigbo et al. 2008) using codon usage information taken from (Nakamura et al. 2000). Each gene from each sequence was calculated individually and the CAI for each gene is given as the mean of all sequences. In codons where one or more bases were ambiguous, the entire codon was removed from the sequence prior to CAI calculation. Where large portions of sequence were missing these strains were excluded.

4.2.3 Copy number

Quantitative PCR was used to estimate copy number in each strain identified as containing the 2µm. qFLP primers (for: TTCAATGAAGGGCCTAACGG and rev: TCGCTCCAATTTCCCACAAC) were used to amplify plasmid DNA and qHO (for: CGGTTTGTGACGATCACGTT and rev: CCTGTTTTCATCTGGACCATCTC) and qpHO (for: GACATCGCAAACGTCACGG and rev: AGAGCCATCTGCGCACATAAC) were used as genomic controls for *S. cerevisiae* and *S. paradoxus* respectively. Each strain was included once per plate for each primer and the plates replicated 3 times, randomizing well positions each time. The two species were amplified on separate plates as they use different genomic primers and therefore required separate standard curves.

4.3 Results and discussion

4.3.1 Prevalence and distribution

A total of 184 *Saccharomyces* strains were screened for the presence of the 2µm plasmid; 68 *S. cerevisiae* and 116 *S. paradoxus* strains. The plasmid was more common among *S. cerevisiae* strains (57.35%, N = 68) than in *S. paradoxus* (23.3%, N = 116). A complete list of all strain and clonal information can be found in table 4.A.1.

Clones

Several of the *S. paradoxus* strains used in this study have been shown to have 100% nucleotide identity across ~4kb at 6 conserved loci, and were therefore identified as being members of the same clone (Johnson et al. 2004; Koufopanou et al. 2006). Among UK strains, 17 of the 26 strains were members of a clone (Johnson et al. 2004), none of which were found to contain the plasmid. Canadian strains contained two clonal groups; four strains were found to be identical to A4, all of which were [cir^0], and 16 strains were identical to A12 (Koufopanou et al. 2006). In the latter group, all clones contained the plasmid, with the exception of A10, which was [cir^0], which indicates that the plasmid has been recently lost in this strain. In addition, Latvian strains SIG1 and SIG8 formed one genotype (Koufopanou et al. 2006), both containing the plasmid.

Geographic distribution

Figure 4.3.1 shows the global distribution of the 2µm in both species. *S. cerevisiae* strains in this collection originate from every major continent (with the exception of Antarctica) and the plasmid was isolated from at least one strain from each region. *S. paradoxus* strains were isolated from three different regions, Europe, North America and the Far East. Plasmid distribution was restricted to two areas, North America and Europe, while no plasmids were found in the 14 Far Eastern stains. North American *S. paradoxus* strains were isolated from one location in Canada. Of these 26 strains, 65.4% contained the plasmid (50% of genotypes (N=8); table 4.3.1). European strains originate from numerous locations within this region. 12.8% contained the plasmid (N=79), however all [cir^+] strains were located in Latvia and Western Russia (figure 4.3.1b), where the plasmid was found in 6/7 and 4/12 strains respectively.

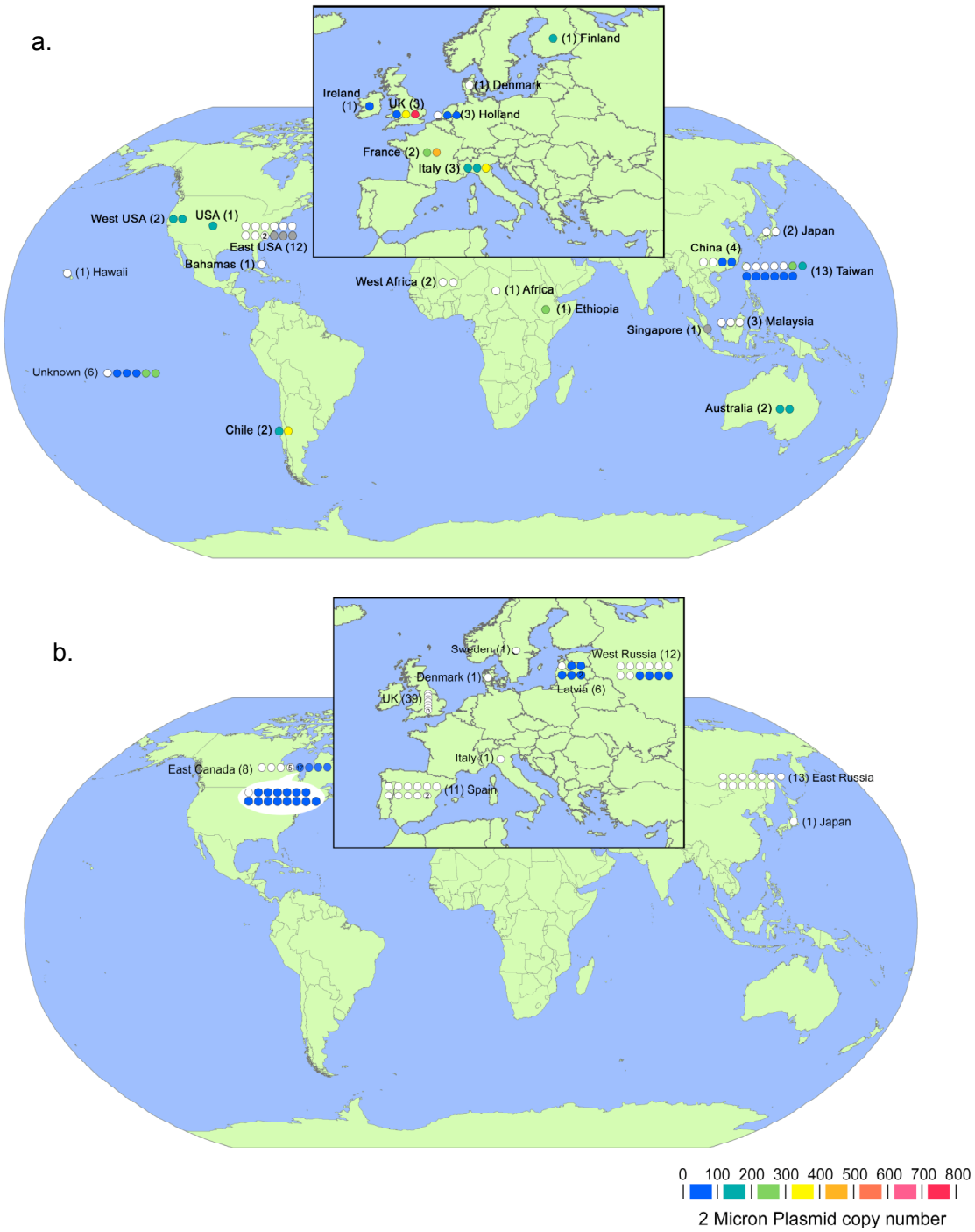


Figure 4.3.1 Geographical distribution of a. *S. cerevisiae* and b. *S. paradoxus* strains. Individual strains are represented by dots. [cir^0] are shown in white, [cir$^+$] strains are coloured with colour representing plasmid copy number as shown in the ledgend. Numbers beside the region names denote the total number of genotypes from each location. Where strains have been identified as members of a clone the clone is represented by a dot with the number of clonal strains shown inside it. In the case of the North American S. *paradoxus* clone A12, strains have also been included in the attached bubble.

Origin and selective history

S. cerevisiae strains have a long history of human association, and can also be catagorised according to their functional origin. Brewing, baking or wine making strains were categorized under Fermentation, while strains isolated from medical patients were classified as Clinical. Finally, strains isolated from wild sources including fruit (except grapes which were assumed to be vineyard associated), oak trees and insects, were considered Wild. The highest prevalence was seen amongst clinical strains, of which 7/7 were [cir$^+$] (table 4.3.1b). 14/20 fermentation strains, and 10/39 of wild isolates contained the plasmid. As *S. paradoxus* strains are not associated with human activities, all strains are classified as wild.

a.

Species	Geographic Origin	N		Percent [cir$^+$]		Copy no.		
		s	g	s	g	mean	s.e	N
S. cerevisiae	Africa	4		25%		229		1
	Australia	2		100%		150	18.5	2
	Europe	13		92.3%		233	61.4	12
	Far East	23		47.8%		60	22.2	11
	North America	17	16	35.3%	43.8%	139	9.3	6
	Oceana	1		0%				
	South America	2		100%		255	81.5	2
	Unknown	6		83.3%		108	61.9	5
S. paradoxus	Europe	76	58	13.2%	15.7%	23	4.6	10
	Far East	14	8	0%				
	North America	26	14	65.4%	49.1%	4	0.7	17

b.

Species	Functional Origin	N		Percent [cir$^+$]		Copy no.		
		s	g	mean	mean	mean	s.e	N
S. cerevisiae	Clinical	7		100%		245	98.2	7
	Fermentation	20		70%		169	36.1	14
	Unknown	2		100%		260	10.3	2
	Wild	39	38	41%	4.47%	74	19.4	17
S. paradoxus	Wild	118	73	23.2%	16.3%	11	2.5	27

Table 4.3.1 Prevalence and copy number of [cir$^+$] strains among yeast from different a. geographical and b. functional origins. The total number of strains (s) and genotypes (g) is shown where different, and the percentage of these that contain the 2µm plasmid. Copy number of [cir$^+$] strains is also shown.

Host phylogeny

Plasmid prevalence is patchy in both species indicating numerous loss/gain events (figure 4.3.2). The phylogenetic relationships between *S. cerevisiae* strains are complex and lack any real global structure (Liti et al. 2009). Several strains form 'clean' lineages, the largest of which is the Wine/European clade, made up of strains either associated with wine fermentation or isolated from Europe. In addition, a number of 'mosaic' strains are shown on long branch lengths in between the Wine/Europe clade and other 'clean' lineages (Liti et al. 2009). These strains share sequence homology with different lineages at different regions of the genome and are likely to be the product of regular recombination between lineages. Maximum divergence between strains of *S. cerevisiae* was estimated as 0.6% (Liti et al. 2009). In addition to these strains, a number of wild strains from the Far East (Taiwan and eastern China) were also surveyed, which were not included in the SGRP and therefore not placed in this phylogeny.

The plasmid was isolated from all ten strains of the Wine/European clade, and in 14/17 mosaic strains. As the latter group all share homology in part with the Wine/European lineage they are likely to be the result of crosses with this lineage, therefore recombination between these strains is an obvious source of plasmid transfer throughout the global *S. cerevisiae* population. Members of the 'clean' Sake, Malaysian and West African lineages were all [cir^0]. In the North American clade, 11 strains provided by P. Sniegowski were not included in the analysis by Liti et al. (2009), but were isolated from the same location (Sniegowski et al. 2002) and therefore likely to belong in this lineage. Three of these strains showed positive PCR results for REP2 and FLP primer pairs, however these results had low repeatability and both sequencing and qPCR failed. Therefore, it is possible that these strains contain a divergent form of the plasmid to which primers to not easily anneal. Alternatively, positive results may simply represent false positives. Therefore the status of plasmids in these strains could not be resolved. The plasmid was identified in 10/15 wild Far Eastern strains.

In *S. paradoxus,* strains from North America, Europe and the Far East form distinct clades (figure 4.3.2b), therefore the pattern of plasmid distribution is the same at that described geographically. Divergence between these lineages has been shown to be between 1.4% (European vs. Far Eastern) and 4% (European/Far Eastern vs. North American) and therefore is far larger than divergence within the *S. cerevisiae* lineages (0 – 0.6%; Liti et al. 2009). Thus, the apparent absence of the plasmid in Far Eastern strains indicates that either it has been lost from this population, or that the plasmids in the European and/or North American populations have been introduced subsequent to this split.

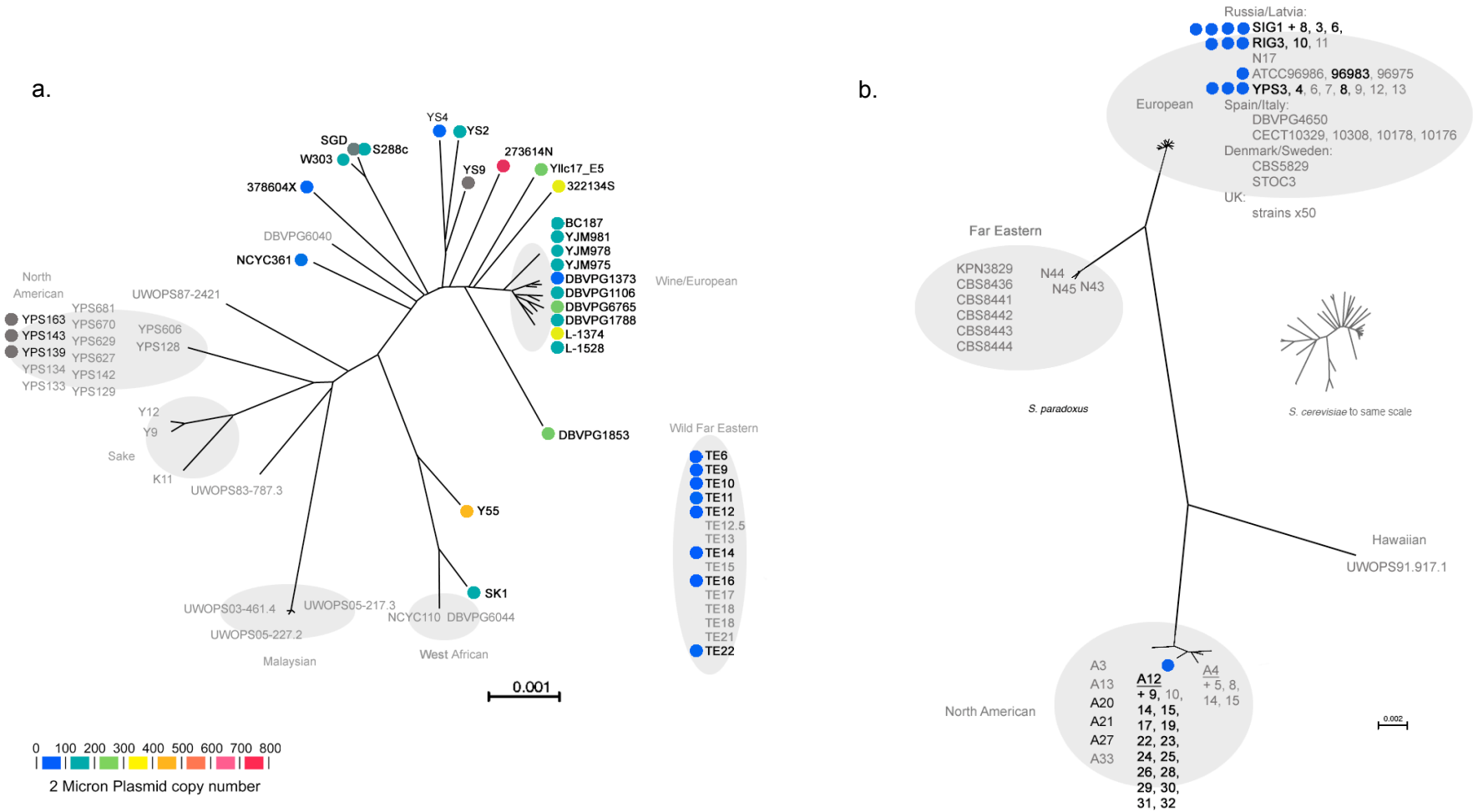


Figure 4.3.2. Prevalence of the 2 micron plasmid among (a.) *S. cerevisiae* and (b.) *S. paradoxus* strains. [cir^0] strains are shown in grey, while strains containing the 2 micron plasmid are black with dots coloured by copy number according to the above scale. In strains for which copy number information was not obtained, dots are shown in grey. Figure adapted from Liti et al (2009).

4.3.2 Copy number variation

Estimates of copy number were variable in both species, with a highly skewed distribution towards low copy number (figure 4.3.3). Variability was most striking in *S. cerevisiae*, where copy number ranged from 5 to 753 (s.e.=0.77, N = 3 and s.e.=40.85, N=3 respectively), with average copy number of ~154 (s.e.=26.47, N=35). In *S. paradoxus* copy number was much lower (mean=11, s.e.=2.52, N=27), ranging from 0.6 to 46 (s.e.=0.046, N=3 and s.e.=5.20, N=3 respectively). RIG10 was the only strain in which copy number appeared to be <1. Sequence data confirm homology to quantitative PCR primers, making this figure unlikely to be due to poor annealing of primers to the plasmid. This result may be due to the partial loss of the plasmid in the culture prior to extraction. If this was the case it would most likely be a consequence of very low copy number in this strain, as this has been shown to affect stability of plasmid (Futcher & Cox 1984).

These extremes in copy number are surprising. The range of plasmid copy numbers, between 2 – 753, amounts to between ~0.05% to 28% of the total nuclear DNA being made up of plasmid DNA. This is likely to make a substantial difference in the cost of carrying the plasmid (*see* Chapter 2). Success of the plasmid is dependent on 3 main

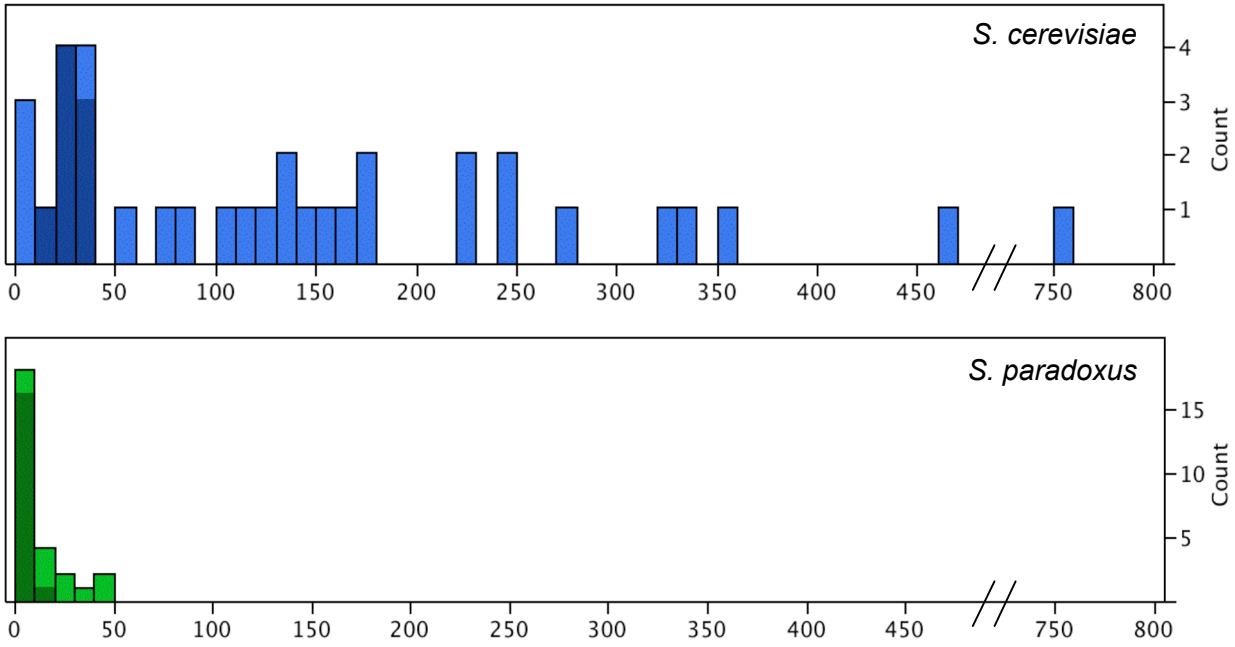


Figure 4.3.3 Copy number distribution among [cir$^+$] strains of *S. cerevisiae* and *S. paradoxus* strains. Strains of wild Far Eastern *S. cerevisiae* are highlighted in dark blue, and strains of North American *S. paradoxus* are highlighted in dark green. Distribution is taken from means of 3 replicate qPCR estimates for each strain. Bin sizes = 10.

factors; the fitness advantage of losing the plasmid, the rate of spontaneous loss of the plasmid and the rate of outcrossing in the host populations (which allows the plasmid to spread through the population; Futcher et al 1988). We have previously shown that plasmid copy number correlates positively with the fitness cost it imposes on the host (*see* Chapter 2). Therefore, at very high copy numbers, we expect the plasmid to be extremely costly, making the fitness advantage of becoming [cir^0] much greater. Conversely, copy number is also associated with stability, as low copy number increases the likelihood of loss through miss-segregation (Futcher et al. 1984). At very low copy numbers, then, the rate of spontaneous plasmid loss in the population is expected to be much higher. Therefore under either very low, or very high, copy numbers we might expect that such extremes would hinder the spread of such plasmids through the population.

In *S. cerevisiae*, copy number appears to be unrelated either to geographic distribution (figure 4.3.1) or phylogeny of the host (figure 4.3.2), with the exception of the wild Far Eastern strains, which all had relatively low copy numbers of between 18 and 39 (s.e.=1.15, N=3 and s.e.=2.77, N = 3 respectively). Conversely, in *S. paradoxus* copy number appears to largely be defined by population. In North American stains, 16 of the 17 strains contained <4 plasmids per cell (mean=3, s.e.=0.13, N=16), with the exception of strain A12 which had an estimated 14 copies (s.e.=0.66, N=3). The disparity between North American and Far Eastern strains is highlighted in figure 4.3.3b. In the European strains copy number ranged from 9 (s.e.=0.24, N=3) to 20 (s.e.=1.39, N=3) in Russian strains, and from 25 (s.e.=2.41, N=3) to 46 (s.e.=5.2, N=3) in Latvian strains, with the exception of RIG10.

4.3.3 DNA Sequence variation

Sequences were obtained for a total of 50 strains, using either the Sanger SGRP Database or PCR based sequencing as described. The full alignment is shown in figure 4.A.2. In all ten Far Eastern *S. cerevisiae* isolates, only two genes were successfully sequenced; FLP and REP2. Outside of these regions, PCR was unsuccessful, suggesting that these plasmids may be too divergent at these sites for our primers to anneal. One strain (UWOPS05-217.3) that had been previously identified as being [cir^0], returned two reads from the BLAST search of SGRP data. This was lower than in other strains (between 13 and 142 reads) and therefore this discrepancy may be the result of contamination. This strain was excluded from the analysis.

Figure 4.3.4 shows that the remaining 39 strains for which the whole sequence was available cluster into 3 distinct clades. Xiao & Rank (1991a) have previously identified two of these clades, as type 1 and 2 plasmids. These can be defined by their sequence

Whole Plasmid

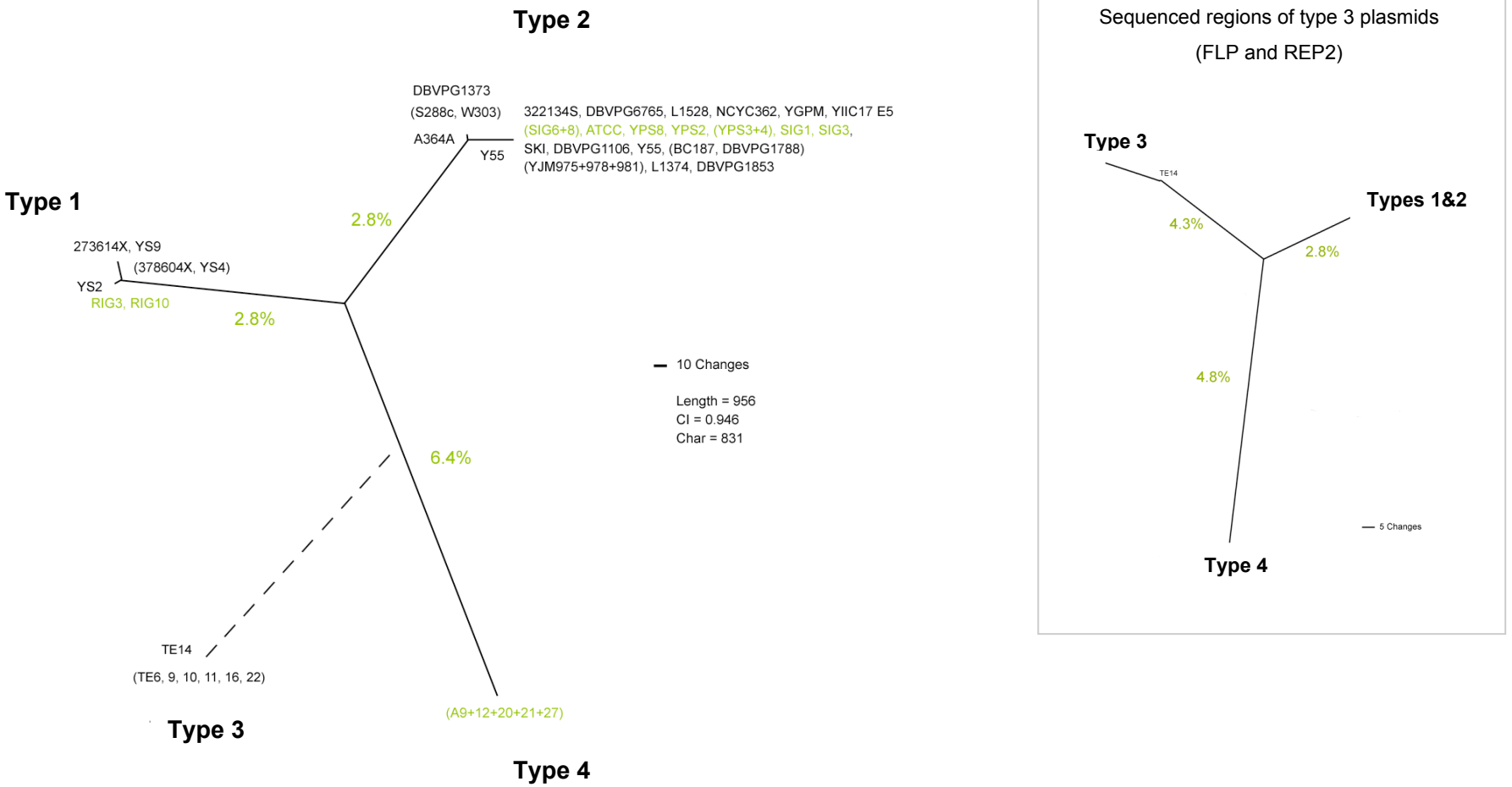


Figure 4.3.4 Maximum parsimony tree for the entire plasmid sequence from strains of *S. cerevisiae* (black) and *S. paradoxus* (green). The plasmids fall into 4 distinct clades. The Far Eastern *S. cerevisiae* plasmids (type 3) were not included in the tree as they were only partly sequenced; the dashed line represents the rough placement of these plasmids on the phylogeny. The inset shows the phylogeny of the four plasmid types at the region sequenced in type 3. The trees were built using a heuristic search, treating gaps as missing. Plasmids identical at every known base connected within parentheses.

at the STB region, which was shown to be ~30% divergent between these two types, and alignment to either type 1 strain YSC7754A (Genbank, M57550) or type 2 strain YSC4110 published STB regions (Genbank, M57549) was used to identify these types. Divergence between these strains across the whole plasmid was ~5.6%. Both type 1 and 2 plasmids were found in *S. cerevisiae* strains as has been described before, but were also found in European *S. paradoxus* (highlighted in figure 4.3.4). Type 1 plasmids were less common and were found in the two Latvian *S. paradoxus* strains and five *S. cerevisiae* mosaic strains. A total of 25 plasmids were identified as type 2, and were found in nine Latvian and Russian *S. paradoxus* strains, and 16 *S. cerevisiae* strains. All 10 Wine/European strains of *S. cerevisiae* contained type 2 plasmids, but they were also found among the mosaic strains.

The presence of both type 1 and 2 plasmids in both *Saccharomyces* species indicates at least two cross-over events have occurred between these species, one for each of the plasmid types. The transfer of nuclear DNA between these populations has been inferred previously. European *S. paradoxus* strains were found to contain a 23kb subtelomeric segment showing >99% identity to *S. cerevisiae*, compared to 85% homology across the rest of the genome (Liti et al. 2006). Liti et al. (2006) suggest that this is the result of introgression, rather than horizontal gene transfer, which is known to be rare in yeast (Liti & Louis 2005). Reproductive isolation between *Saccharomyces* spp. has been demonstrated, but is not hard and fast (Greig 2009), and *S. cerevisiae* and *S. paradoxus* are known to coexist within a few centimeters of each other in the wild (Sniegowski et al. 2002). Therefore, matings between these species are also likely to be a means of 2µm plasmid transfer.

In addition to the type 1 and 2 plasmids described by Xiao & Rank (1991), two novel genotypes were identified in the wild Far Eastern *S. cerevisiae* strains, and in the North American *S. paradoxus*. Both are unique to their respective populations and appear to be highly divergent, therefore, in accordance with the nomenclature of these plasmids we defined the Far Eastern *S. cerevisiae* plasmids as type 3 and North American *S. paradoxus*, type 4. As only 2 genes were partially sequenced in the type 3 plasmids these were not included in the main phylogeny (however figure 4.3.4 shows the phylogeny based on this region of the plasmids only). Type 3 plasmids are highly divergent from the types 1 and 2 plasmids (7.1%) and also the type 4 plasmids (9.1%) for this region. Type 4 plasmids, which were entirely sequenced, are ~6.3% divergent from the types 1 and 2 plasmids.

Two strains contained plasmids identical to the published 2µm sequence, A364A, which has been previously identified as a hybrid (Xiao & Rank 1996). These plasmids align predominately with the type 2 plasmids, but share ~160bp of sequence identical to the

type 1 plasmids in the region directly adjacent to the FRT site on the right hand side of the plasmid.

Plasmid structure

All plasmids showed the same basic structure described previously, and genes and *cis*-acting regions were identified from previous annotations. Some small variations in the number of repeating units and the length of some genes were observed, shown in figure 4.3.5. Previous work by Xiao et al. (1991a) showed variation at a number of repeating regions among the type 2 plasmids. We found that most type 2 sequences had two repeats of the 125bp STB region, with the exception of strain DBVPG1373, which had one copy. We also found one strain (SIG1) with a 31bp tandem repeat in the inverted repeat region, although all other plasmids had just one copy.

An additional repeating region was found in the inverted repeats of the type 1 and 2 plasmids. These plasmids contain a TATA region on the LUR of variable length; 12, 24 or 34bp. All three variants were found in the type 2 plasmids and the 12 and 34bp variants found in type 1 plasmids. Type 4 plasmids did not contain any such TATA region.

Discontinuity between Left and Right Unique Regions

Type 1 and 2 plasmids have previously been shown to contain identical LURs, while the RUR is relatively divergent (Xie et al. 1994). Figure 4.3.6 shows the phylogenetic structure when each side is analysed separately. No fixed differences were found between the type 1 and 2 plasmids in the LUR, supporting the finding of Xie et al. (1994). Between type 1 & 2 and type 4, divergence was 7.3% within the LUR. Conversely at the RUR, all three types appear to have diverged to a similar extent (~11.4% between type 1 and type 2 and 5.4% between types 1 & 2 and type 4; figure 4.3.6). This is comparable to divergence between the two host species, *S. cerevisiae* and *S. paradoxus*, ~10 – 15% (Kellis et al. 2003; Liti et al. 2009). Divergence at the LUR is slightly lower, however still remains larger than maximum divergence with the *S. cerevisiae* (0.6%) and *S. paradoxus* populations (4%).

The LUR has previously been reported as being identical among the type 1 and 2 plasmids based on RFLP analysis (Xie et al. 1994). However, we have identified a small amount of sequence variation in this region in both plasmid types (π x100 = 0.06 and 0.03 in type 1 and 2 respectively; table 4.3.3). Comparisons of the phylogeny of the LUR and RUR regions demonstrate that the phylogenetic topology of these regions is incompatible. While the type 1 and 2 plasmids form monophyletic groups at the RUR,

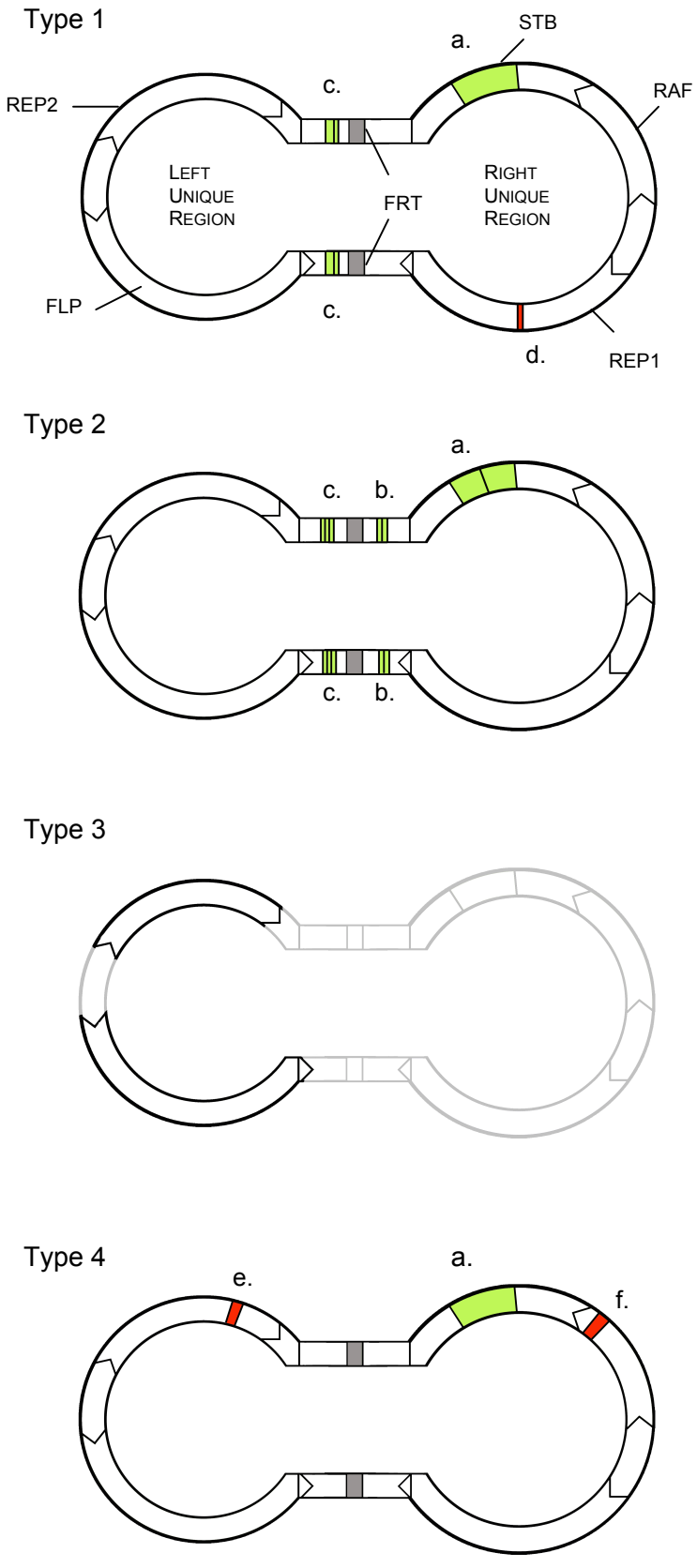


Figure 4.3.5 Major features of the 2 micron plasmid seen in this study.

All plasmids show the same basic structure. Type 3 plasmids were only sequenced at the FLP and REP2 loci. Sequencing failed for the remainder of the plasmid (outlined in grey). Sites with structural variation are highlighted in green. Deletions in coding regions are shown in red.

a. STB regions of Type 2 plasmids consist of 1 (N=1) or 2 (N=25) copies of a 125bp repeat sequence. Type 1 and 4 plasmids have no repeats in this region.

b. 31bp tandem repeat in one strain

c. Variable TATA region of 12, 24 and 34bp. All found in type 2. 10 and 34bp variants in type 1. Absent from type 4.

d. 6bp/2 amino acid deletion REP1

e. 36bp/6 amino acid deletion FLP

f. 7bp frameshift deletion. Alters the last 5 amino acids in RAF protein, however protein length preserved as new stop codon is inserted at the same position.

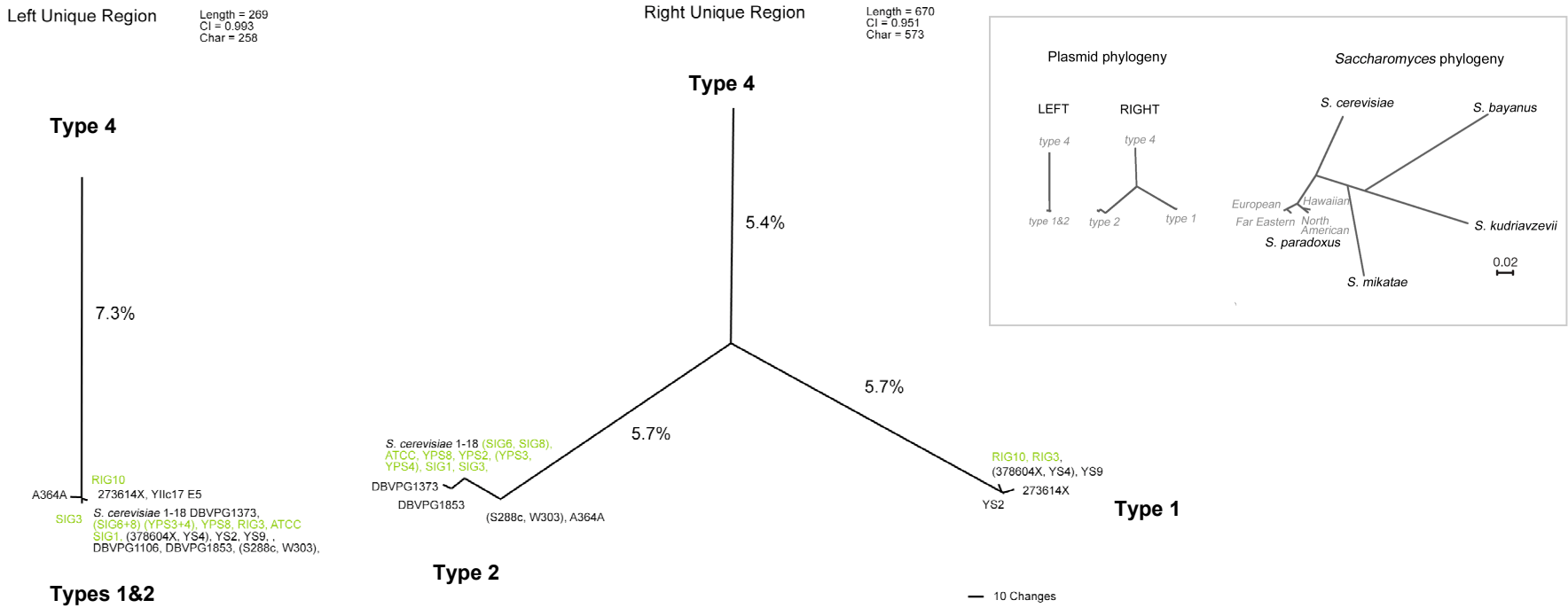


Figure 4.3.6 Parsimony tree for the two halves of the plasmid, divided from the center of the two FRT sites of recombination. The Far Eastern S. cerevisiae plasmids were excluded as sequences were not complete. The tree was built using a heuristic search, treating gaps as missing. Plasmids identical at every known base connected within parentheses. Inset shows trees to scale with a phylogeny of *Saccharomyces* yeasts adapted from Liti et al (2009).

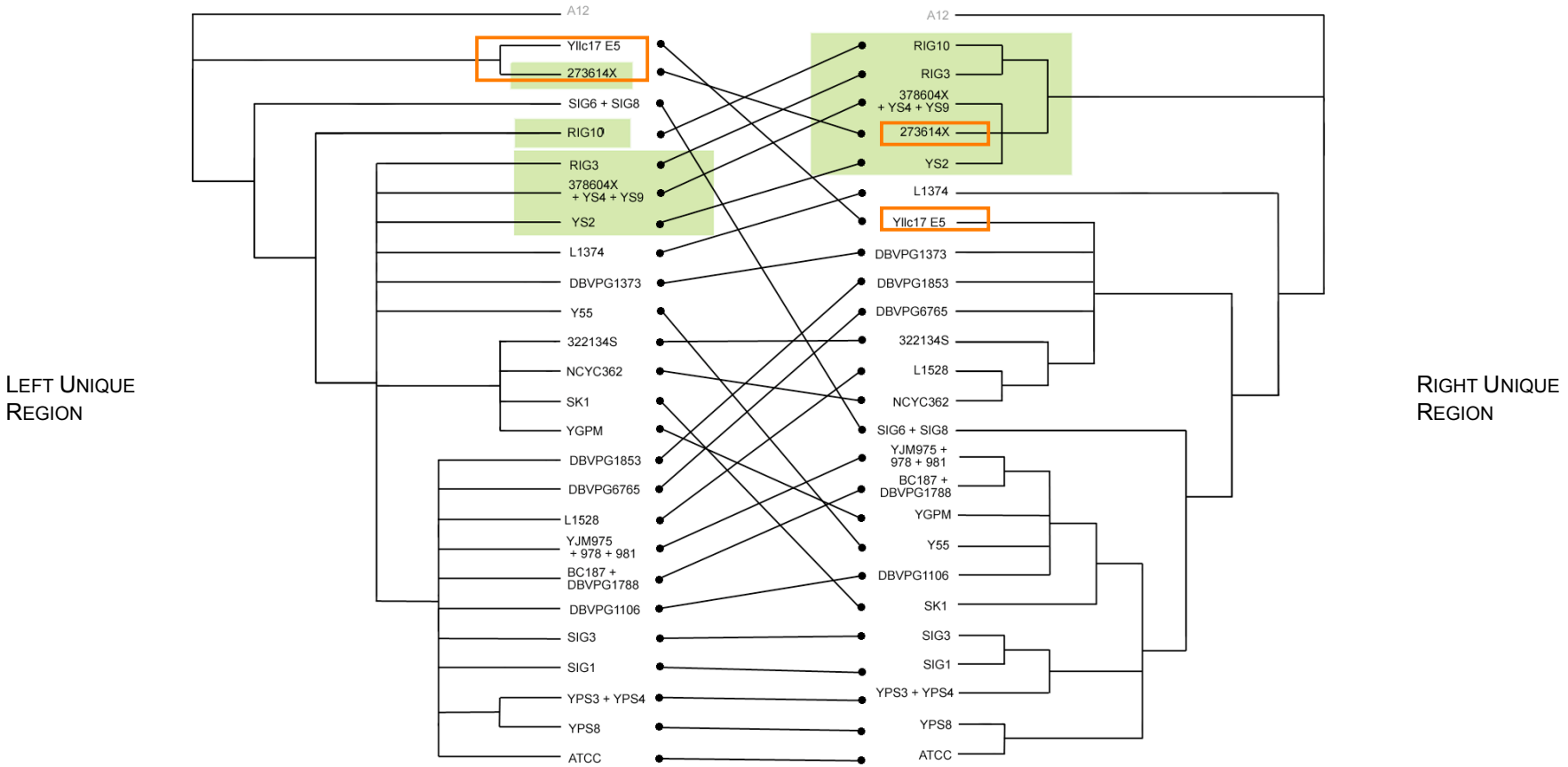


Figure 4.3.7 Cladograms of the left and Right Unique Regions demonstrating incompatibility between the two sides of the type 1 and 2 plasmids. Parsimony trees constructed by heuristic search of 100 trees, treating gaps as missing and using strain A12 as an outgroup. At the Right Unique Region (RUR), the type 1 plasmids (green) form a monophyletic group, however this is not the case in the Left Unique Region (LUR). Strains circled in red give an example of strains with identical LURs, which are divergent from other strains, but divergent RURs.

they do not at the LUR (figure 4.3.7). The best example of this is the plasmids from *S. cerevisiae* strains 273614X and YIIc17 E5 (circled, figure 4.3.7). These plasmids both contain the most divergent LUR compared to other type 1 and 2 plasmid, with 8 SNPs and a 32bp TATA box not seen in any other strains. However at the RUR these plasmids fall into the type 1 and 2 clades respectively. The incompatibility in topologies of the LUR and RUR suggest that the two sides of the plasmid can be exchanged between different plasmids (in effect, a plasmid form of sex), and may explain the extreme conservation of the LUR relative to the divergent RUR. This may be the consequence of two possible mechanisms, bias gene conversion, suggested by Xie et al (1994) and site-specific transfer, which we propose here. The two mechanisms are illustrated in figure 4.3.8.

Bias gene conversion: Xie et al (1994) suggest that gene conversion occurs between the FRT sites of the plasmid, which is both directional, so that it is the LUR and not the RUR that is converted, and polarised, so that one LUR (i.e. the one we observe in populations) always converts the other LUR variant. This would imply that the LUR is effectively able to drive, by 'overwriting' other LUR variants from the population. Gene conversion mediated by the 2μm FLP-FRT has been seen previously in chromosomal constructs (Rank et al. 1988). However, the mechanism for gene conversion in this case is not fully understood (Xie et al 1994).

Site-specific recombination: A second, and possibly more parsimonious explanation for this observed pattern is that the two sides of the plasmid are able to recombine. Plasmid copy number is controlled by site-specific recombination by the plasmid-encoded recombinase, Flpp, at the two plasmid FRT sites. This allows one unique region to be inverted relative to the other (see figure 1.3 chapter 1). However it is also possible that recombination occurs between the FRT sites of different plasmids, rather than between the two FRT sites on one genome. This has been evoked previously as a possible explanation for the hybrid nature of the published 2μm sequence (Xiao & Rank 1996). Figure 4.3.8 shows how recombination between two plasmids can be resolved either into the two original forms or into two hybrid forms, therefore creating four possible variants in one population. If this occurred between the ancestral type 1 and 2 plasmids, the loss of one of the LUR may have been the result of a selective sweep removing one variant from the population.

As site-specific recombination is already mechanism known to be used by the plasmid, the latter explanation is possibly more likely. In all plasmids sequenced at this region, the FRT site is 100% conserved, suggesting that recombination between different plasmids would be possible. These two hypotheses could be tested by combining marked plasmids into single hosts. Under the former hypothesis one LUR variant would

be lost from the population, whist under the latter, all unique regions would persist but alter their combinations in the host.

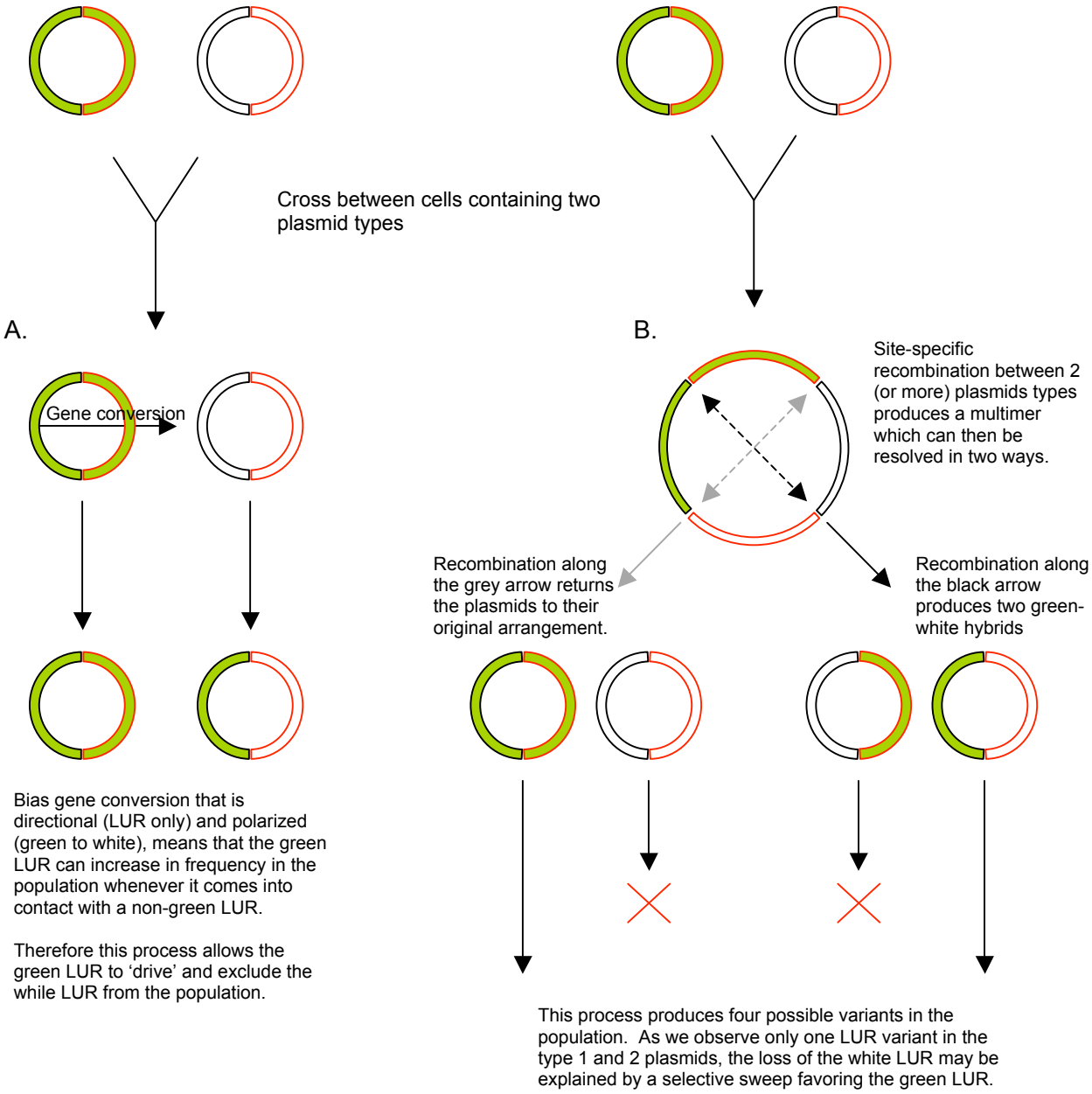


Figure 4.3.8 Putative routes for homogenization of the Left Unique Region (LUR), which is found to be identical between types 1 and 2, while the Right Unique Region (RUR) is divergent. The LUR is shown outlined in black and the RUR is outlined in red. Green and white plasmids are combined into the same cell during a cross between different strains. The loss of the white LUR may occur through a. bias gene conversion, or b. site specific recombination followed by the loss of variants carrying the is LUR.

It has previously been suggested that the divergence in the RAF, REP1 and STB loci (located on the RUR) of the type 1 and 2 plasmids has been driven by coevolution between the two ORFs, and the *cis* acting STB locus (Xiao et al 1991a). However we note that, under either hypothesis discussed above, this pattern of divergence is explained by the different selection on the LUR and RUR region; whereby the LUR becomes conserved, either through gene conversion or through selection, but the RUR remains divergent. Therefore the divergence seen in the RAF, REP1 and STB loci can be most simply explained by their shared presence on the RUR (i.e. the linkage between them), and although it is possible that these loci have been driven by coevolution, linkage may provide a simpler explanation for this pattern.

Polymorphism and divergence

A number of plasmids appeared to have identical genotypes (shown in figure 4.3.4). In total five pairs, one triplet and all five type 4 plasmids were identical at every known base. For the type 4 plasmids this is unsurprising as host strains were highly related and two were members of a clone. One pair and a triplet were found in strains from the Wine/European *S. cerevisiae* clade, two in the mosaic strains of *S. cerevisiae,* including the hybrid plasmids which are found in strains S288c and W303. The final two pairs were found among the European *S. paradoxus*. Interestingly, one was isolated from SIG8 and SIG6, despite the fact that the strain SIG8 was identified as a clone of SIG1. In fact the SIG1 plasmid differs from the latter by 5 SNPs and has different numbers of the TATA and tandem repeat regions. This discrepancy may be due to these strains not being entirely clonal. Alternatively, this pattern could be observed in clone mates where two plasmids co-infect a clone, but different plasmid genotypes have come to dominate in different clonal cell lines.

Similar levels of polymorphism were observed in type 1 and 2 plasmids over the entire sequence (π x100 =0.033 and 0.026 respectively; table 4.3.2), however the pattern of polymorphism differs across the genome (figure 4.3.8). Among type 2 plasmids polymorphism is highest in the STB region (π x100 =0.085) compared to an average of 0.045 in the remaining non-coding regions. Conversely, type 1 plasmids contained no polymorphism the majority of the RUR, including the STB region, but had much higher diversity in the LUR compared to type 2 plasmids (figure 4.3.8). Polymorphism was also high in the IG2 region (the intergenic region between REP2 and FLP; π x100 =0.101), which has no known function. In both types, little to no variation was observed in the region from REP1 to RAF.

As has been described previously, divergence between type 1 and 2 plasmids was restricted to the RUR only. Divergence was highest in the STB region ~21% compared

		Left Unique Region (LUR)						Right Unique Region (RUR)					LUR	RUR	WHOLE
		IG1	**REP2**	IG2	**FLP**	IG3		**REP1**	IG4	**RAF**	IG1				
		IR1 left				IR2 left	IR2 right				STB	IR1 right			
Type 1 N=7	length (bp)	356	**891**	372	**1272**	190	177	**546**	261	**1118**	665	234	3081	3001	6082
	S	0	**0**	3	**7**	1	1	**0**	0	**0**	0	1	11	2	13
	NS	0	**1**	0	**0**	0	0	**0**	0	**0**	0	0	1	0	1
	dS (x 100)	0	**0**	0.4	**0.27**	0.26	0.28	**0**	0	**0**	0	0.21	0.2	0.037	0.12
	dN (x 100)	0	**0.019**	0	**0**	0	0	**0**	0	**0**	0	0	0.0062	0	0.0032
	π (x 100)	0	**0.014**	0.101	**0.069**	0.066	0.071	**0**	0	**0**	0	0.053	0.0556	0.0095	0.033
Type 2 N=26	Length* (bp)	-	**-**	-	**-**	-	-	**-**	262	**1124**	903	239	-	3251	6332
	S	2	**0**	3	**6**	2	1	**2**	0	**2**	20	1	13	24	37
	NS	0	**2**	0	**2**	0	0	**2**	0	**2**	0	0	4	2	6
	dS (x 100)	0.085	**0**	0.12	**0.071**	0.16	0.085	**0.027**	0	**0.027**	0.33	0.063	0.064	0.11	0.088
	dN (x 100)	0	**0.012**	0	**0.0081**	0	0	**0.0092**	0	**0.0092**	0	0	0.0067	0.0032	0.0049
	π (x 100)	0.022	**0.0086**	0.031	**0.024**	0.041	0.022	**0.014**	0	**0.014**	0.085	0.016	0.021	0.031	0.026
Type 3 N=8 (REP2) N=3 (FLP)	Length* (bp)		**480**		**514**								994		994
	S		**0**		**1**								1		1
	NS		**0**		**0**								0		0
	dS (x 100)		**0**		**0.25**								0.049		0.049
	dN (x 100)		**0**		**0**								0		0
	π (x 100)		**0**		**0.065**								0.013		0.013
Type 4 N=5	Length* (bp)	359	**1272**	-	**855**	178	185	**1124**	263	**539**	717	242	3036	3070	6106
	S	0	**0**	0	**0**	0	0	**0**	0	**0**	0	0	0	0	0
	NS	0	**0**	0	**0**	0	0	**0**	0	**0**	0	0	0	0	0

** where different from type 1*

Table 4.3.2 Polymorphisms within plasmid types across the 2μm genome, divided genic (bold) and intergenic regions (IG). The Inverted Repeats (IR) are divided by the FRT site into left and right, depending on which unique region they belong. Data shown are the number of synonymous (S) and nonsynonymous (NS) polymorphisms, S per synonymous site (ds), NS per nonsynonymous site and overall polymorphism per site (π). Values are multiplied by 100 where shown.

Plasmid type		Left Unique Region (LUR)						Right Unique Region (RUR)					LUR	RUR	WHOLE
		IG1	**REP2**	IG2	**FLP**	IG3		**REP1**	IG4	**RAF**	IG1				
		IR1 left				IR2 left	IR2 right				STB	IR1 right			
Type 1 vs 2	Length (bp)	356	**891**	372	**1272**	190	177	**546**	261	**1118**	685	233	3081	3020	6101
	S	0	**0**	0	**0**	0	22	**23**	28	**52**	143	28	0	296	296
	NS	0	**0**	0	**0**	0	0	**20**	0	**24**	0	0	0	44	44
	dS (x 100)	0	**0**	0	**0**	0	0.49	**0.17**	0.42	**0.18**	0.82	0.47	0	0.38	0.19
	dN (x 100)	0	**0**	0	**0**	0	0	**0.049**	0	**0.029**	0	0	0	0.02	0.0097
	D	0%	**0%**	0%	**0%**	0%	12.4%	**7.9%**	10.7%	**6.8%**	20.9%	12%	0	11.3%	5.57%
Type 3 vs 4	Length (bp)		**480**		**514**								994		994
	S		**28**		**21**								52		52
	NS		**23**		**17**								38		38
	dS (x 100)		**0.23**		**0.19**								0.21		0.21
	dN (x 100)		**0.064**		**0.04**								0.05		0.05
	D		**10.6%**		**7.8%**								9.1%		9.1%
Type 1/2 vs 4	Length (bp)	343	**873**	371	**1272**	165	176	**539**	261	**1118**	620	231	3024	2945	5969
	S	24	**41**	42	**48**	7	9	**11**	13	**24**	64	13	162	134	296
	NS	0	**38**	0	**22**	0	0	**13**	0	**13**	0	0	60	26	86
	dS (x 100)	0.27	**0.18**	0.44	**0.15**	0.17	0.2	**0.08**	0.2	**0.084**	0.41	0.22	0.21	0.18	0.19
	dN (x 100)	0	**0.058**	0	**0.023**	0	0	**0.032**	0	**0.016**	0	0	0.027	0.012	0.019
	D	7%	**9.1%**	11.3%	**5.5%**	4.2%	5.1%	**4.5%**	5%	**3.3%**	10.3%	5.6%	7.3%	5.4%	6.4%
Type 1/2 vs 3/4	Length (bp)		**480**		**514**								994		994
	S		**5**		**11**								14		14
	NS		**12**		**2**								14		14
	dS (x 100)		**0.041**		**0.19**								0.055		0.21
	dN (x 100)		**0.034**		**0.04**								0.019		0.05
	D		**3.5%**		**2.5%**								2.8%		2.8%

Table 4.3.3 Sequence divergence between plasmid types across the 2μm genome, labeling as shown in table 4.3.2. Data shown are the number of synonymous (S) and nonsynonymous (NS) fixed differences between types, S per synonymous site (dS), NS per nonsynonymous site (dN) and overall divergence per site (D).

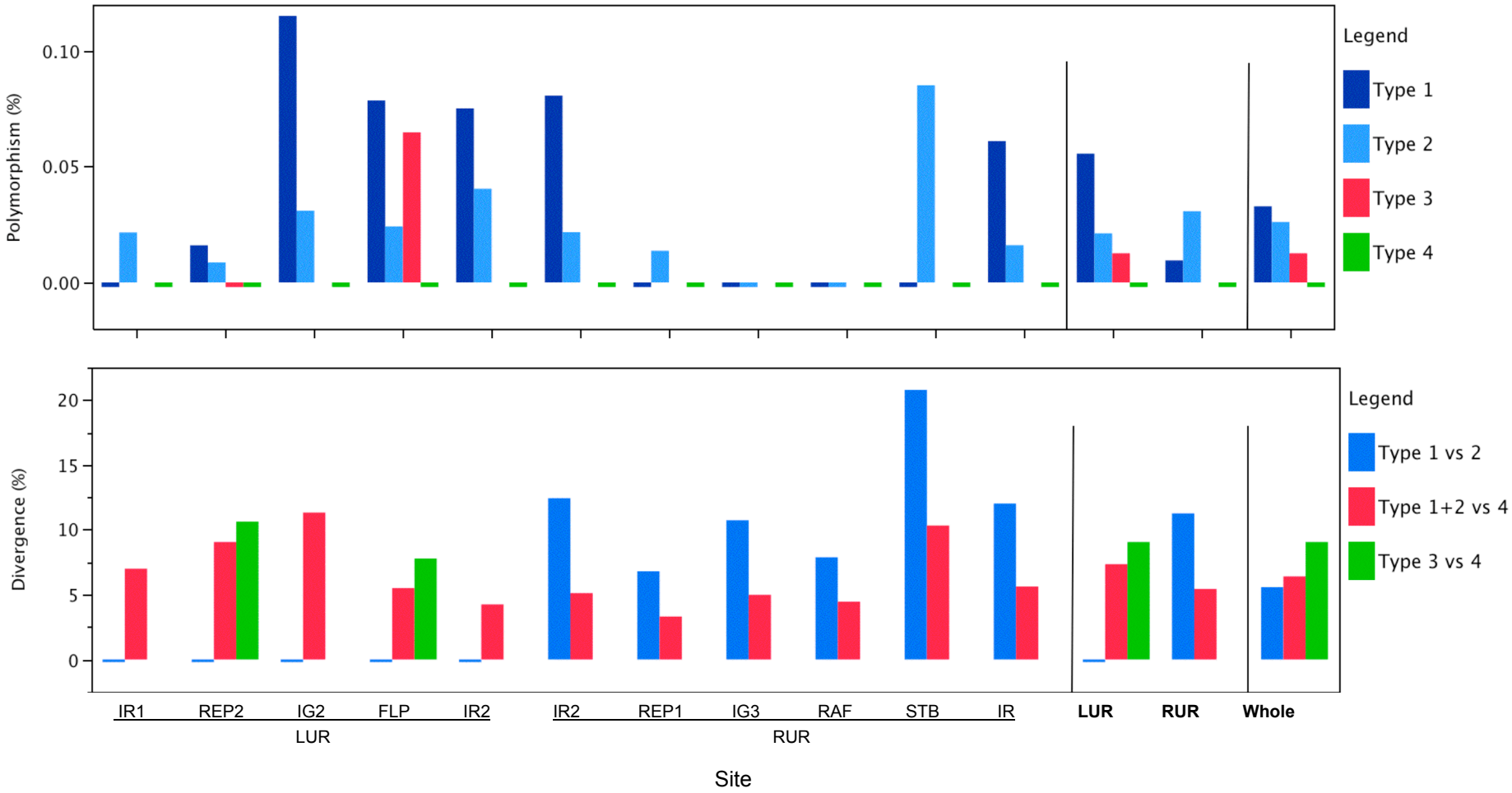


Figure 4.3.8 Divergence and polymorphism across the length of the plasmid genome, divided into the major plasmid loci. Labeling as shown in table 4.3.2. Bars at zero indicate zero polymorphism/divergence, while no bars indicate no data for this region.

Type 1 vs 2:

LUR

	S	NS	Total
Polymorphic	24	5	**29**
Fixed	0	0	**0**
Total	**24**	**5**	**397**

na

RUR

	S	NS	Total
Polymorphic	26	2	**57**
Fixed	296	44	**340**
Total	**346**	**51**	**397**

χ^2=0.795, df= 1, *p*=0.37

Whole Plasmid

	S	NS	Total
Polymorphic	50	7	**57**
Fixed	296	44	**340**
Total	**346**	**51**	**397**

χ^2=0.019, df= 1, *p*=0.89

Type (1&2) vs 4:

LUR

	S	NS	Total
Polymorphic	24	5	**29**
Fixed	162	60	**222**
Total	**186**	**65**	**251**

χ^2=1.280, df= 1, *p*=0.26

RUR

	S	NS	Total
Polymorphic	322	46	**368**
Fixed	134	26	**160**
Total	**456**	**72**	**528**

χ^2=1.332, df= 1, *p*=0.25

Whole Plasmid

	S	NS	Total
Polymorphic	346	51	**397**
Fixed	296	86	**382**
Total	**642**	**137**	**779**

χ^2=12.552, df= 1, *p*=0.0004*

Table 4.3.4 2x2 contingency tables for to test for selection in, (a.) type 1 vs type 2 plasmids and, (b.) type 1 and 2 vs type 4 plasmids (Macdonald & Kretiman 1991). Nucleotide variation was too low to test each gene individually. Therefore tests were performed on the LUR, RUR and whole plasmid genome as shown. The number of synonymous (S) and nonsynonymous (NS) mutations calculated as the total number of polymorphic sites in one population plus the comparative population (polymorphic) and as sites with fixed differences between the two populations (fixed)

to ~12% at other intergenic regions on this half of the plasmid (table 4.3.3, figure 4.3.8b). Divergence at the STB was also high between the type 4 and types 1 & 2 plasmids, ~10.3% compared to ~5% at intergenic regions on the right side of the plasmid. However, divergence was equally high (11.3%) at IG2 on the LUR.

High polymorphism and divergence at the STB locus suggests that this region is evolving faster than other loci. High divergence in STB has been previously demonstrated in the types 1 and 2 plasmids by Xiao et al. (1991). The STB region acts as a centromere for the plasmid molecule, and it has been suggested that the point centromeres of Sacharromycete yeasts are in fact derived from the ancestral 2μm (Malik & Henikoff 2009) (for discussion of this see Chapter 1). Recent analysis of centromere diversity in *S. paradoxus* suggests that centromeric regions are evolving much faster than neutrally evolving sites on the chromosome (Bensasson et al. 2008). This does not appear to be as a result of selection however, as polymorphism is also high in this region. Bensasson et al (2008) suggest that the source of this high mutation rate may be increased exposure to mutagenic conditions, for instance by the physical bending involved in wrapping around histones. The functional similarity of the STB region and chromosomal centromeres would suggest that this region might also be exposed to similar mutagentic conditions.

The lack of polymorphism in the region from REP1 to RAF may indicate strong purifying selection acting at this site. Due to the low amount of polymorphisms in this data set, it was difficult to test for selection within these genomes. However, this was possible at the unique region and whole plasmid levels, using 2x2 contingency tables (table 4.3.5) to test for differences in the ratio of synomymous and nonsynonymous differences within and between populations (McDonald & Kreitman 1991). No significant deviation from neutrality was detected between type 1 and 2 plasmids, including the RUR (in which the REP1 – RAF region is located). Between the type 1 & 2 and type 4 plasmids (note that the type 1 and 2 plasmids can also be considered as one recombining population) we detected purifying selection on the whole plasmid (χ^2=12.552, df= 1, p=0.0004), but not at each side independently.

Plasmid divergence compared to host strains

In order to more accurately estimate divergence of the plasmid relative to the host genes we calculated codon usage bias for all four plasmid genes using the Codon Adaptation Index (CAI), (Sharp & Li 1987). Codon usage bias represents the extent to which certain codons are used preferentially over others encoding the same amino acid (Sharp et al. 2010). As the genetic code is degenerate, multiple codons can code for one amino acid, therefore the use of these different codons would, under the simplest

hypothesis, be random. The CAI is a measure of the deviation of a sequence from this random sampling. CAI was calculated for each sequence and averaged to produce one overall value for each gene (*as* Hirsh et al 2005). All four of the genes showed high codon bias (FLP=0.77, RAF=0.748, REP1=0.743 and REP2=0.759) and were significantly higher than the average CAI value for nuclear genes (mean CAI=0.300, wilcoxon sign-rank$_{3039}$=-231, p<.0001, N=3036; nuclear gene CAI data from Hirsh et al 2005). Bias among nuclear genes ranges from 0.145 to 0.892, with a strongly skewed distribution towards lower CAI values. As a result, all four plasmid genes fall into the top 1% of yeast ORFs (figure 4.3.9).

The high codon bias observed in the plasmid ORFs may act to constrain the rate of evolution between plasmid types (Hirsh et al 2005). Figure 4.3.9 shows a comparison between synonymous divergence (dS) in genes from each plasmid type and between *S. cerevisiae* and *S. paradoxus* nuclear genes. In nearly all comparisons shown, divergence between the different plasmid types, at each gene, is extremely close to the divergence predicted for a nuclear gene in the host species' (figure 4.3.9). Two exceptions to this are observed. The type 1 and 2 LUR ORFs are entirely conserved

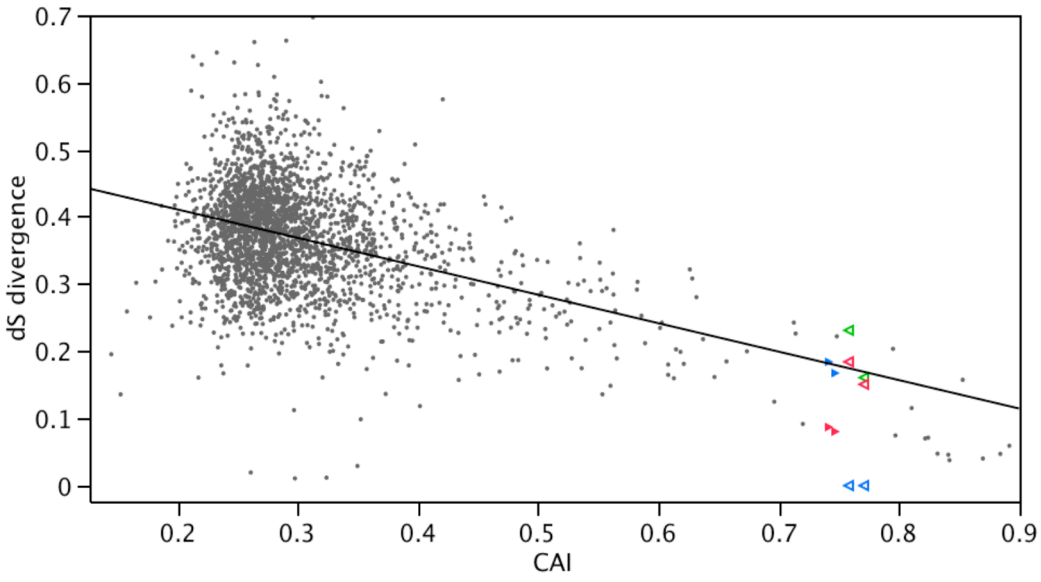
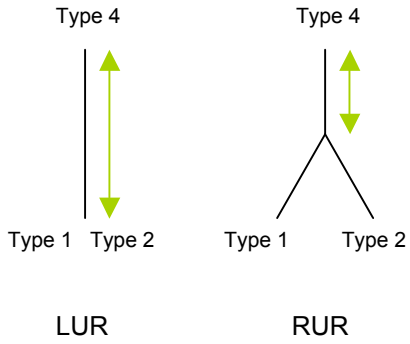


Figure 4.3.9 Synonymous divergence (dS) between *S. cerevisiae* and S. *paradoxus* at 5237 ORFs, compared to codon usage bias (CAI) for each plasmid gene. For each plasmid gene, dS was calculated for type 1 vs 2 (blue), type 3 vs 4 (green) types 1 & 2 vs 4 (red). Genes on the LUR (FLP and REP2) are shown by open arrows, and on the RUR (REP1 and RAF) by closed arrows. Data for dS of nuclear genes provided by M. Hahn (pers. comm). Data for CAI of nuclear genes taken from Hirsh et al. (2005).

Figure 4.3.10 Explanation of the divergence between types 1 & 2 plasmids from type 4. As the LUR of the types 1 and 2 are identical, divergence between these and the type 4 plasmids represents the distance between two branch tips (illustrated by green arrows). At the RUR divergence is more similar to the distance between type 4 and TMCRA of types 1 and 2.



(dS=0), which has previously been discussed. In addition, divergence between the RUR of types 1 & 2 and type 4 is around half that predicted by this slope (RAF=0.08 and REP1=0.084 compared to predicted values of 0.181 and 0.184 based on the slope dS = 0.493 - 0.417 * CAI), while at the LUR divergence was much more similar to predicted values. However this is expected, as this comparison analogous to the divergence between type 4 and TMCRA of types 1 and 2 (described in figure 4.3.10).

This suggests that divergence between all four plasmid types is approximately equal to that expected in nuclear genes, with equally high codon bias, between the two host species. This is perhaps surprising as, in the case of the type 1 and 2, plasmids are found circulating in the same population, and from the existence of hybrid plasmids and evidence for possible recombination, we can infer that they are sometimes combined in the same host cell. This may imply that these plasmids have evolved in isolation and only more recently, perhaps through human trafficking, have been brought into contact with each other.

High codon usage bias in the 2µm

The high codon usage bias of the 2µm plasmid genes may also be, itself, indicative of the evolution of this element. Codon bias is largely associated with genes which are highly expressed in the cell (Tuller et al. 2010), either through selection of increasing translational efficiency (Tuller et al. 2010) or increasing translational accuracy (Akashi 1994). We have previously shown that plasmid DNA appears to be 'cheaper' per kb, in terms of fitness cost, than nuclear genes (*see* chapter 2). The majority of the cost associated with additional DNA in the cell appears to be due to the translation of proteins (Stoebel et al. 2008; Torres et al. 2007), specifically by reducing ribosome and tRNA availability (Stoebel et al. 2008). Therefore, increased codon bias may represent one way of minimising the cost of plasmid carriage, as optimal codons can be quicker to process (Sorensen & Pedersen 1991). As we have argued in chapter 2, multi-copy

elements, such as the 2μm, are likely to be subject to selection to reduce the cost of DNA to the cell from both a within-individual and between-individual selection perspective.

Elements that have a transmission advantage by being numerous will be under selection to maximise their copy number to increase transmission (within-individual selection), which is balanced against between-individual selection, which acts to minimise the degree of harm inflicted on the individual. Therefore selection will favour any adaptation to reduce the cost of the element to the individual, but without reducing transmission. The high copy number of the element is analogous to gene expression, as the degree of selection is amplified by the amount of protein being produced (Tuller et al. 2010). Thus, theoretically, this would be the same, whether the protein is expressed at high levels from one gene, or at regular levels from many copies.

4.5 Conclusions

This study has demonstrated that the 2μm plasmid is a common, but not universal element is *Saccharomyces* spp. populations, which represent anything from just 0.04%, all the way up to a quarter of the nuclear DNA of the host. The plasmid can be found in both *S. cerevisiae* strains, of all geographic and functional origins, and also in the wild species *S. paradoxus*. The distribution of plasmid is largely explained by the host population structure, which is to be expected as the element's sole mode of transmission is vertical. Thus, plasmid relies on sexual reproduction, within and between yeast populations, to spread.

We have also demonstrated several features of the plasmids life history not previously identified:

Firstly, plasmids have, on at least two occasions, been transferred between the two species. This exchange appears to be somewhat geographically isolated as both events are restricted to *S. paradoxus* strains from North Eastern Europe. Matings between *S. cerevisiae* strains and the European *S. paradoxus* populations have been previously inferred by Liti & Louis (2005), as these strains share a region of localized high nucleotide identity, possibly as a result of introgression. Interestingly, this signal is observed in the genomes of *S. paradoxus* from across Europe, while the 2μm is found only in strains from Russia and Latvia, and not from western European strains. Therefore, if the transfer of the plasmid also occurred during this mating event, we may assume that the plasmid has been lost from populations surveyed from the west of Europe but retained in the north-eastern populations. This could be due to population

effects such as a higher rate of outcrossing, or genomic factors such as strong resistance to the plasmid in western European *S. paradoxus*. Alternatively, this pattern of plasmid distribution may be due to crosses between this *S. paradoxus* population, and *S. cerevisiae* strains occurring more frequently in this location. However, our sampling of European *S. paradoxus* strains is far from comprehensive and does not include strains from central Europe. An interesting extension of this would, therefore, be to sample wild strains more broadly across the continent to examine the real extent and frequency of the plasmid's distribution.

Secondly, the incompatibility of the LUR and RUR regions of the type 1 and 2 plasmids indicates that recombination between the two sides of the plasmid is occurring in these mixed populations. We have speculated that this has lead to the loss of one form of the plasmid LUR from either the type 1 or 2 plasmids, by allowing one variant to spread to fixation independently of its RUR. We also note that the assumption that recombination occurs between the LUR and RURs of different plasmids is analogous to sexual recombination between chromosomes, and that this results in the genes of either unique region being connected by linkage. Therefore the divergence of the RUR loci; RAF, REP1 and STB, in these plasmid types can be most simply explained by linkage, and does not necessarily indicate functional coevolution between the two genes and the STB region, which has been previously proposed to explain this pattern of divergence (Xiao et al 1991a). Indeed, we may expect that the REP2 gene would be a stronger candidate for coevolution with REP1 and STB, than RAF, as the REP proteins form a complex that interacts with the STB region to ensure the stable inheritance of the plasmid (*see* Chapter 1). Therefore it is interesting that the fixation of one LUR in this population has occurred, despite one member of this protein complex, REP2, being on this region, while the REP1 ORF and STB locus are located on the divergent RUR region.

We have also found that the copy number of this 'high copy number plasmid' can in fact be extremely variable, including copy numbers as low as 2, and as high as 753. This suggests a huge difference in the selective pressures on copy number in different populations. We have shown in Chapter 3 that one factor which can influence the evolution of copy number is the outcrossing rate of the population. One potential hypothesis for the disparity in copy numbers is a difference is rates of sex, with low copy number plasmids found in more prudish populations and high, variable copy numbers in more frequently sexual populations. As outcrossed sex determines the rate the 2μm can spread through a population, plasmids in rarely sexual populations will be subject to strong selection against being costly, selecting for low copy number elements. In more sexual populations, selection is weaker as the plasmid is able to drive, therefore within-individual selection drives the evolution of high copy numbers. As a consequence, selection for either resistance (reducing copy number) or tolerance (permitting high copy

number) would also be high, potentially creating a mixture of high and low copy numbers dependant on host and plasmid genotypes.

Alternatively, the control of copy number may entirely be due to the hosts level of control over copy number. In this case, populations such as the North American *S. paradoxus* may be extremely resistant to plasmids increasing in copy number. These competing hypothesizes could be tested by the factorial exchange of plasmids into different host backgrounds, to determine to what extent the host and plasmid define copy number.

Finally, sequence analysis has confirmed the finding of Xiao et al (1991b), who found that the variation in nucleotide divergence was particularly high in the STB region of the plasmid. This indicates that this region is evolving much faster the majority of the genome, and is analogous to the fast rate of evolution at the centromeres of yeast chromosomes (Bensasson et al. 2008). We have also identified particularly high codon bias at all ORFs in the different plasmid genome. This, in addition to the finding that plasmid DNA appears to decrease host fitness at a lower rate per kb than chromosomal DNA (*see* Chapter 2), suggests that the plasmid is under strong selection to reduce the cost of its gene expression and suggests that selfish genetic elements, particularly, those for whom high copy number is adaptive, would be subject to this selection, comparable to long or highly expressed genes (Akashi 2003).

To our knowledge, this study is the first to examine of the population biology of the 2µm plasmid on a global scale, and in particular in the wild species, *S. paradoxus*. Although the plasmid is largely defined by its host populations, its movement between populations, and even species, demonstrates that it should be considered an element in its own right. Its movement between populations also represents a useful signal for inferring matings between populations, and therefore this, and other elements like it, may prove a useful indicator for the study of their host populations.

4.A Appendix

Table 4.A.1 Complete list of strains used in this study. Footnote denote strains that have been found to be likely members of a clone, based on homology across ~4000bp sequenced from 6 loci (Johnson et al. 2004; Koufopanou et al. 2006). 2 micron plasmid presence (1) or absence (0) was determined by a combination of PCR and sequencing, and the copy number determined by qPCR. Numbers to the left of certain strains are used in reference to figure 4.3.6.

		Origin		Source			
Spp.	Strain	Continent	Area	Type	2µm	Copy No. (s.e.)	
Saccharomyces cerevisiae							
	Y12	Africa	Africa	Fermentation, Palm wine	0		
	NCYC110	Africa	Africa, West	Fermentation, Ginger Beer	0		
	DBVPG6044	Africa	Africa, West	Fermentation, Wine	0		
	DBVPG1853	Africa	Ethiopia	Wild, Plant	1	229.36	(31.69)
	YS2	Australia	Australia	Fermentation, Baking	1	168.56	(23.85)
1	DBVPG1106	Australia	Australia	Wine	1	131.50	(23.38)
2	DBVPG1788	Europe	Finland	Wild, Soil	1	177.79	(23.97)
3	YIIc17	Europe	France	Fermentation, Wine	1	242.64	(31.10)
4	Y55	Europe	France	Fermentation, Wine	1	466.52	(95.15)
5	DBVPG6040	Europe	Holland	Wild, Fruit	0		
	DBVPG1373	Europe	Holland	Wild	1	70.99	(11.45)
6	NCYC361	Europe	Ireland	Fermentation, Beer	1	38.70	(5.48)
7	YJM981	Europe	Italy	Clinical	1	106.48	(6.12)
8	YJM978	Europe	Italy	Clinical	1	354.32	(51.31)
9	YJM975	Europe	Italy	Clinical	1	119.95	(12.57)
	YS4	Europe	Holland	Fermentation, Baking	1	85.49	(13.51)
	378604X	Europe	UK	Clinical	1	52.90	(1.62)
10	322134S	Europe	UK	Clinical	1	322.31	(39.12)
	273614X	Europe	UK	Clinical	1	752.86	(40.85)
	TE16	Far East	China	Wild, Fruit	1	17.91	(1.15)
	TE15	Far East	China	Wild, Fruit	0		
	TE14	Far East	China	Wild, Fruit	1	21.06	(1.21)
	TE13	Far East	China	Wild, Fruit	0		
	Y9	Far East	Japan	Fermentation, Sake	0		
	K11	Far East	Japan	Fermentation, Sake	0		
	UWOPS05-227.2	Far East	Malaysia	Wild, Insect	0		
	UWOPS05-217.3	Far East	Malaysia	Wild, Plant	0		
	UWOPS03-461.4	Far East	Malaysia	Wild, Plant	0		
	YS9	Far East	Singapore	Fermentation, Baking	1	–	
	TE6	Far East	Taiwan	Wild, Insect	1	22.99	(4.31)
	TE7	Far East	Taiwan	Fermentation, Wine	1	144.23	(1.46)
	TE8	Far East	Taiwan	Fermentation, Wine	1	229.53	(4.68)
	TE9	Far East	Taiwan	Wild, Insect	1	29.32	(0.49)
	TE10	Far East	Taiwan	Wild, Insect	1	31.66	(3.42)

	TE11	Far East	Taiwan	Wild, Insect	1	39.07	(2.77)
	TE12	Far East	Taiwan	Wild, Insect	1	32.79	(6.58)
	TE12.5	Far East	Taiwan	Wild, Insect	0		
	TE18	Far East	Taiwan	Wild, Resin	0		
	TE19	Far East	Taiwan	Wild, Resin	0		
	TE21	Far East	Taiwan	Wild, Bark	0		
	TE22	Far East	Taiwan	Wild, Bark	1	26.93	(0.18)
	TE17	Far East	Taiwan	Wild, Resin	0		
	UWOPS83-787.3	North America	Bahamas	Wild, Fruit	0		
11	SK1	North America	USA	Wild, Soil	1	131.82	(25.17)
	YPS 681	North America	USA, East	Wild, Oak	0		
	YPS 670	North America	USA, East	Wild, Oak	0		
	YPS 629	North America	USA, East	Wild, Oak	0		
	YPS 627	North America	USA, East	Wild, Oak	0		
	YPS 606[111]	North America	USA, East	Wild, Oak	0		
	YPS 142	North America	USA, East	Wild, Oak	0		
	YPS 129	North America	USA, East	Wild, Oak	0		
12	YPS 163	North America	USA, East	Wild, Soil	1	–	
13	YPS 143	North America	USA, East	Wild, Soil	1	–	
14	YPS 139	North America	USA, East	Wild, Soil	1	–	
	YPS 134	North America	USA, East	Wild, Soil	0		
	YPS 133	North America	USA, East	Wild, Soil	0		
	YPS 128	North America	USA, East	Wild, Soil	0	–	
15	BC187	North America	USA, West	Fermentation, Wine	1	157.51	(22.29)
	S288c	North America	USA, West	Wild, Fig	1	127.65	(18.06)
	UWOPS87-2421	Oceana	Hawaii	Wild, Plant	0		
16	L1528	South America	Chile	Fermentation, Wine	1	173.90	(21.13)
17	L1374	South America	Chile	Fermentation, Wine	1	336.85	(12.53)
	BCRC 20270	Unknown		Clinical	1	5.42	(0.77)
	SC3	Unknown		Fermentation, Baking	1	7.09	(1.07)
	SC2	Unknown		Fermentation, Baking	1	9.98	(0.88)
	SC5	Unknown		Fermentation, Baking	0		
	W303	Unknown	Unknown	Unknown	1	249.51	(16.88)
18	DBVPG6765	Unknown	Unknown	Unknown	1	270.14	(7.31)
Saccharomyces paradoxus							
	CBS 5829	Europe	Denmark	Wild, Soil	0		
	DBVPG4650	Europe	Italy	Wild, Guano	0		
	SIG 1	Europe	Latvia	Wild	1	25.41	(2.41)
	SIG 8[112]	Europe	Latvia	Wild	1	31.15	(0.84)
	SIG 3	Europe	Latvia	Wild	1	43.13	(3.12)
	SIG 6	Europe	Latvia	Wild	1	46.44	(5.20)
	RIG 3	Europe	Latvia	Wild	1	26.53	(0.61)
	RIG 10	Europe	Latvia	Wild	1	0.56	(0.05)
	RIG 11	Europe	Latvia	Wild	0		
	N17	Europe	Russia, West	Wild, Oak	0		
	ATCC 96975	Europe	Russia, West	Wild, Soil	0		
	ATCC 96983	Europe	Russia, West	Wild, Oak	1	11.62	(0.22)
	ATCC 96986	Europe	Russia, West	Wild, Oak	0		
	YPS 3	Europe	Russia, West	Wild	1	19.86	(1.39)
	YPS 4	Europe	Russia, West	Wild	1	17.56	(0.72)
	YPS 6	Europe	Russia, West	Wild	0		
	YPS 7	Europe	Russia, West	Wild	0		
	YPS 8	Europe	Russia, West	Wild	1	8.71	(0.24)
	YPS 9	Europe	Russia, West	Wild	0		
	YPS 12	Europe	Russia, West	Wild	0		

[111] Spore isolate of **YPS 142** (Sniegowski et al 2002)
[112] Putative clone of **SIG1***

YPS 13	Europe	Russia, West	Wild	0
CECT10176	Europe	Spain	Wild	0
CECT10178	Europe	Spain	Wild	0
CECT10308	Europe	Spain	Wild, Oak	0
CECT10329[113]	Europe	Spain	Wild, Oak	0
STOC 3	Europe	Sweden	Wild	0
Z1	Europe	UK	Wild, Oak	0
Z1.1	Europe	UK	Wild, Oak	0
Y1	Europe	UK	Wild, Oak	0
Y2	Europe	UK	Wild, Oak	0
Y2.2	Europe	UK	Wild, Oak	0
Y2.8[114]	Europe	UK	Wild, Oak	0
Y3[115]	Europe	UK	Wild, Oak	0
Y4	Europe	UK	Wild, Oak	0
Y4.5	Europe	UK	Wild, Oak	0
Y5	Europe	UK	Wild, Oak	0
Y5.1[116]	Europe	UK	Wild, Oak	0
Y5.6[6]	Europe	UK	Wild, Oak	0
Y6	Europe	UK	Wild, Oak	0
Y6.5	Europe	UK	Wild, Oak	0
Y7	Europe	UK	Wild, Oak	0
Y7.1[117]	Europe	UK	Wild, Oak	0
Y8.1[6]	Europe	UK	Wild, Oak	0
Y8.4[5]	Europe	UK	Wild, Oak	0
Y8.5[4]	Europe	UK	Wild, Oak	0
Y8.6	Europe	UK	Wild, Oak	0
Y8.8[118]	Europe	UK	Wild, Oak	0
Y9[119]	Europe	UK	Wild, Oak	0
Y9.5[9]	Europe	UK	Wild, Oak	0
Y10[120]	Europe	UK	Wild, Oak	0
W7	Europe	UK	Wild, Oak	0
T8.1	Europe	UK	Wild, Oak	0
T18.2	Europe	UK	Wild, Oak	0
T21.4[6]	Europe	UK	Wild, Oak	0
T26.3	Europe	UK	Wild, Oak	0
T46	Europe	UK	Wild, Oak	0
T62.1	Europe	UK	Wild, Oak	0
T68.1	Europe	UK	Wild, Oak	0
T68.2[5]	Europe	UK	Wild, Oak	0
T76.6	Europe	UK	Wild, Oak	0
S36.7	Europe	UK	Wild, Oak	0
Q6.1	Europe	UK	Wild, Oak	0
Q4.1	Europe	UK	Wild, Oak	0
Q14.4	Europe	UK	Wild, Oak	0
Q15.1	Europe	UK	Wild, Oak	0
Q16.1[121]	Europe	UK	Wild, Oak	0
Q31.4[122]	Europe	UK	Wild, Oak	0
Q32.3	Europe	UK	Wild, Oak	0
Q43.5[10]	Europe	UK	Wild, Oak	0
Q59.1	Europe	UK	Wild, Oak	0
Q62.5	Europe	UK	Wild, Oak	0

[113] Putative clone of **CECT10329***
[114] Putative clone of **Y2.2***
[115] Putative clone of **Y1***
[116] Putative clone of **Y4***
[117] Putative clone of **Y5***
[118] Putative clone of **Y6.5***
[119] Putative clone of **Y8.6***
[120] Putative clone of **Y2***
[121] Putative clone of **Q15.1***
[122] Putative clone of **T76.6***

Q69.8[123]	Europe	UK	Wild, Oak	0		
Q70.8[5]	Europe	UK	Wild, Oak	0		
Q74.4[10]	Europe	UK	Wild, Oak	0		
Q89.8	Europe	UK	Wild, Oak	0		
Q95.3[124]	Europe	UK	Wild, Oak	0		
IFO1804	Far East	Japan	Wild, Oak	0		
KPN3829	Far East	Russia, East	Wild, Oak	0		
CBS8436	Far East	Russia, East	Wild, Oak	0		
N43	Far East	Russia, East	Wild, Oak	0		
N44	Far East	Russia, East	Wild, Oak	0		
N45	Far East	Russia, East	Wild, Oak	0		
CBS8440	Far East	Russia, East	Wild, Oak	0		
CBS8441	Far East	Russia, East	Wild, Oak	0		
CBS8442	Far East	Russia, East	Wild, Oak	0		
CBS8443	Far East	Russia, East	Wild, Oak	0		
CBS8444	Far East	Russia, East	Wild, Oak	0		
A3	North America	Canada, East	Wild	0		
A4	North America	Canada, East	Wild, Oak	0		
A5[125]	North America	Canada, East	Wild, Soil	0		
A8[15]	North America	Canada, East	Wild	0		
A15[15]	North America	Canada, East	Wild, Soil	0		
A14[15]	North America	Canada, East	Wild, Soil	0		
A9[126]	North America	Canada, East	Wild	1	2.50	(0.03)
A10[16]	North America	Canada, East	Wild	0		
A12	North America	Canada, East	Wild, Soil	1	13.74	(0.66)
A17[16]	North America	Canada, East	Wild, Soil	1	2.92	(0.24)
A19[16]	North America	Canada, East	Wild	1	3.73	(0.36)
A22[16]	North America	Canada, East	Wild	1	3.57	(0.37)
A23[16]	North America	Canada, East	Wild	1	2.63	(0.08)
A24[16]	North America	Canada, East	Wild, Soil	1	2.64	(0.30)
A25[16]	North America	Canada, East	Wild, Soil	1	2.24	(0.08)
A26[16]	North America	Canada, East	Wild	1	2.88	(0.05)
A28[16]	North America	Canada, East	Wild	1	3.11	(0.25)
A29[16]	North America	Canada, East	Wild	1	2.04	(0.41)
A30[16]	North America	Canada, East	Wild	1	3.70	(0.19)
A31[16]	North America	Canada, East	Wild	1	2.56	(0.21)
A32[16]	North America	Canada, East	Wild	1	2.09	(0.16)
A13	North America	Canada, East	Wild, Soil	0		
A20	North America	Canada, East	Wild	1	2.86	(0.30)
A21	North America	Canada, East	Wild, Soil	1	2.65	(0.16)
A27	North America	Canada, East	Wild, Soil	1	3.28	(0.44)
A33	North America	Canada, East	Wild, Soil	0		

[123] Putative clone of **Q14.4***
[124] Putative clone of **Z1***
[125] Putative clone of **A4***
[126] Putative clone of **A12***

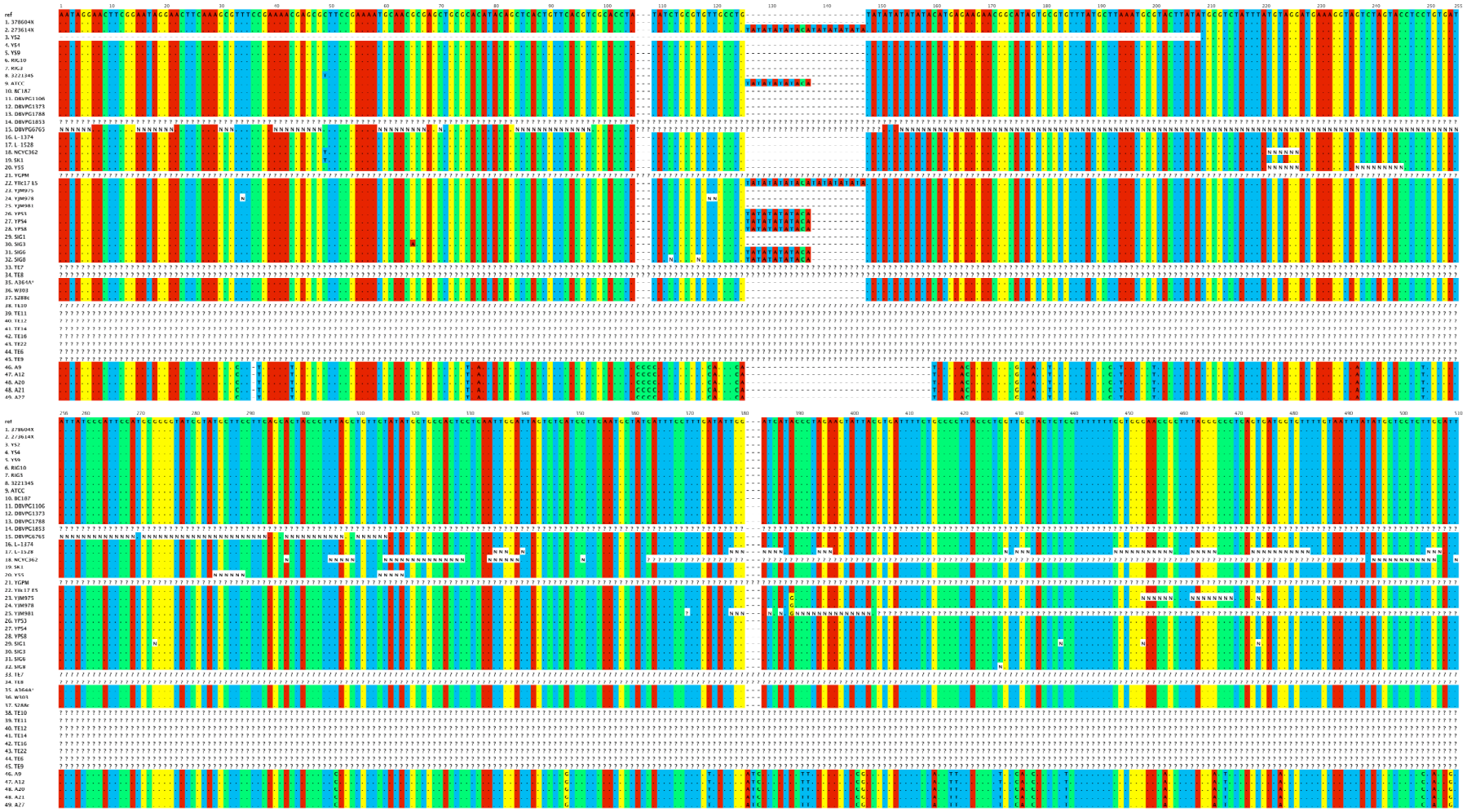
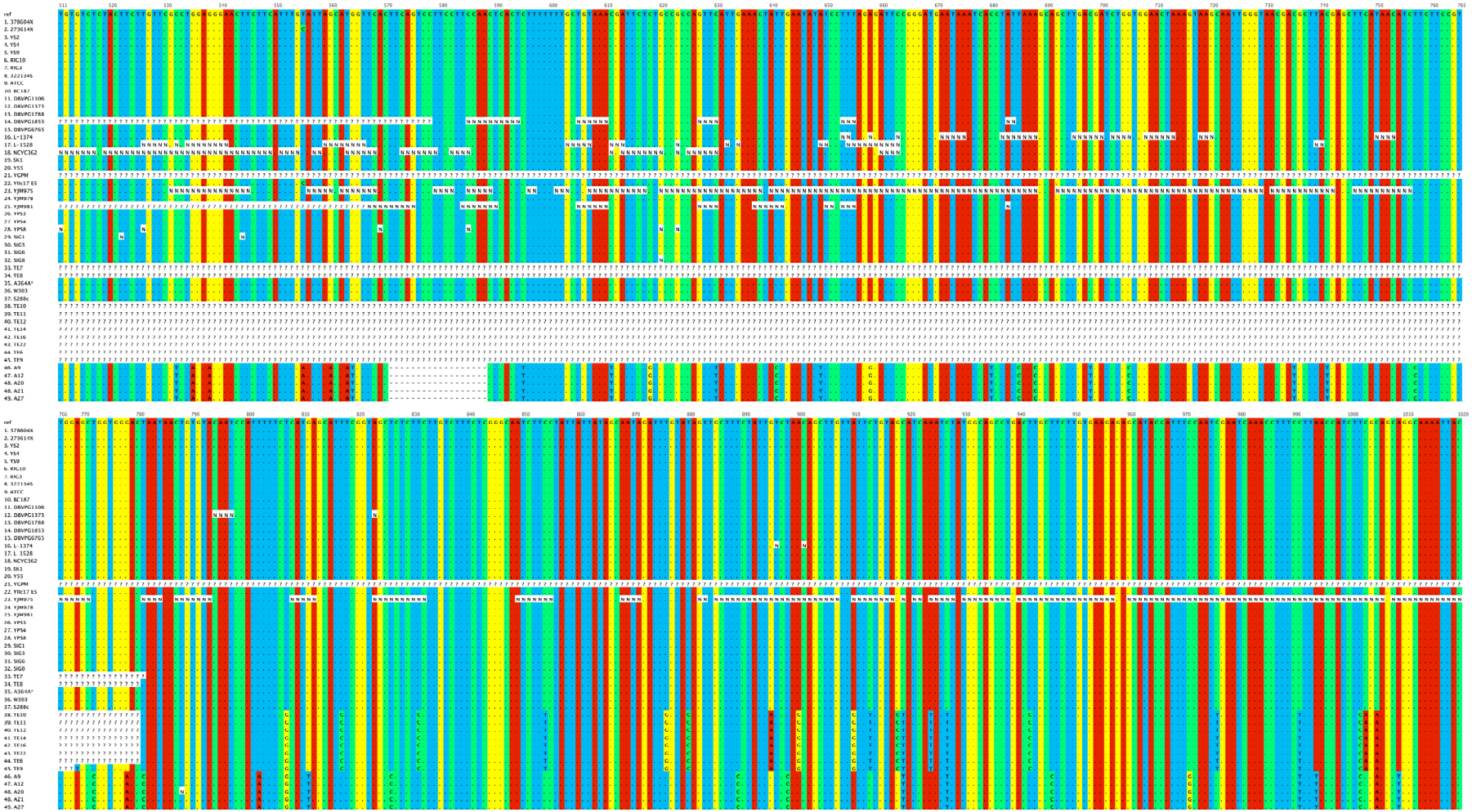
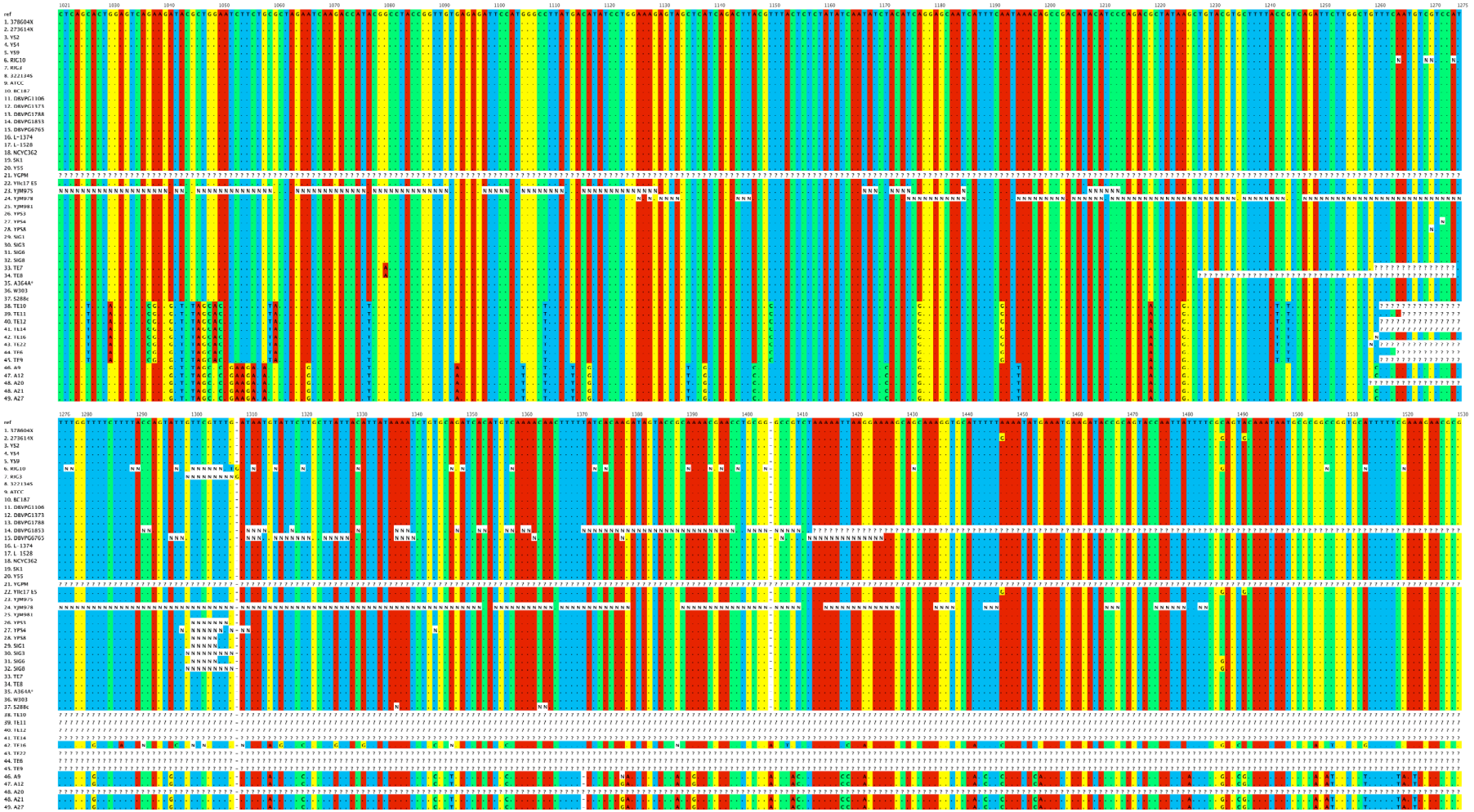
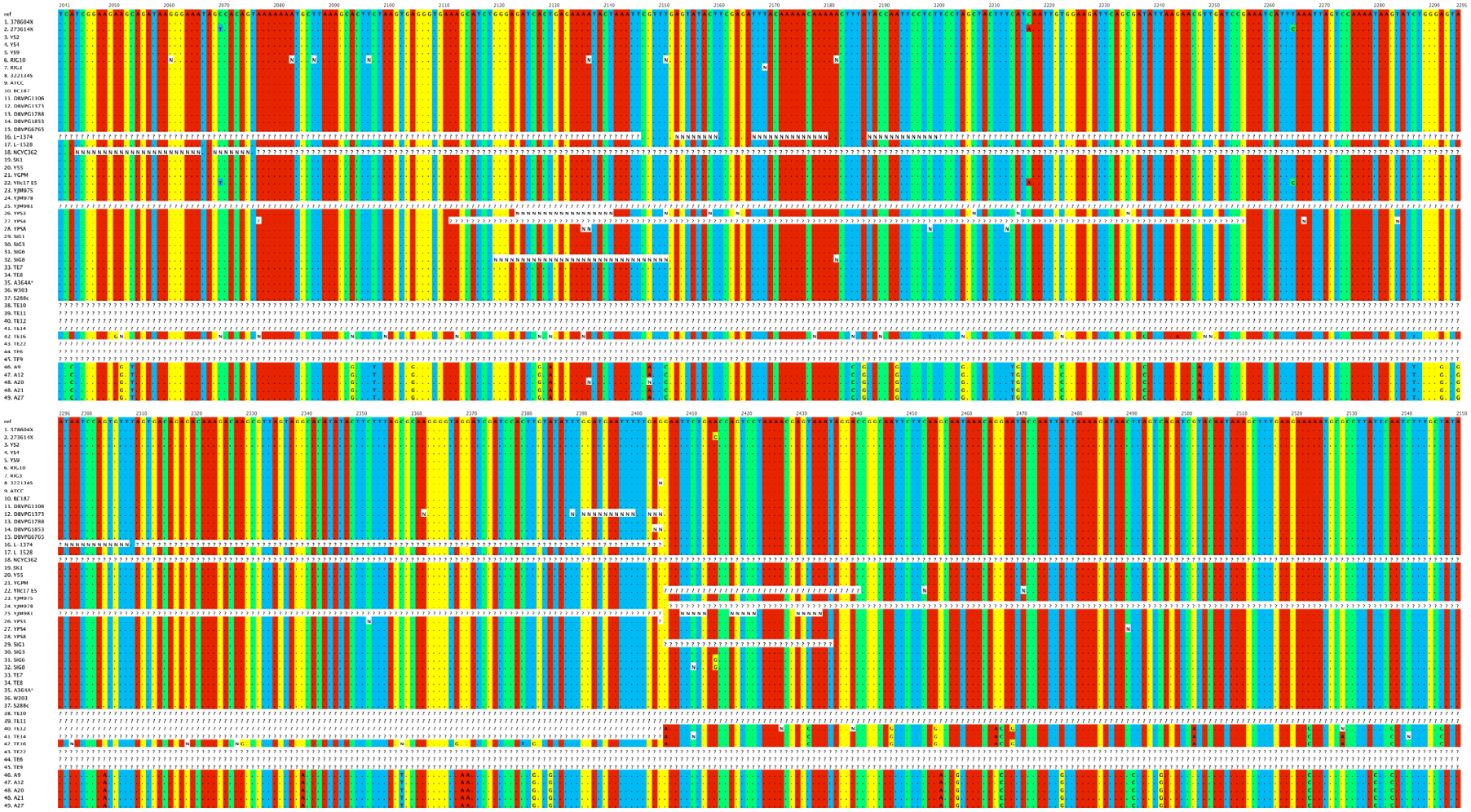
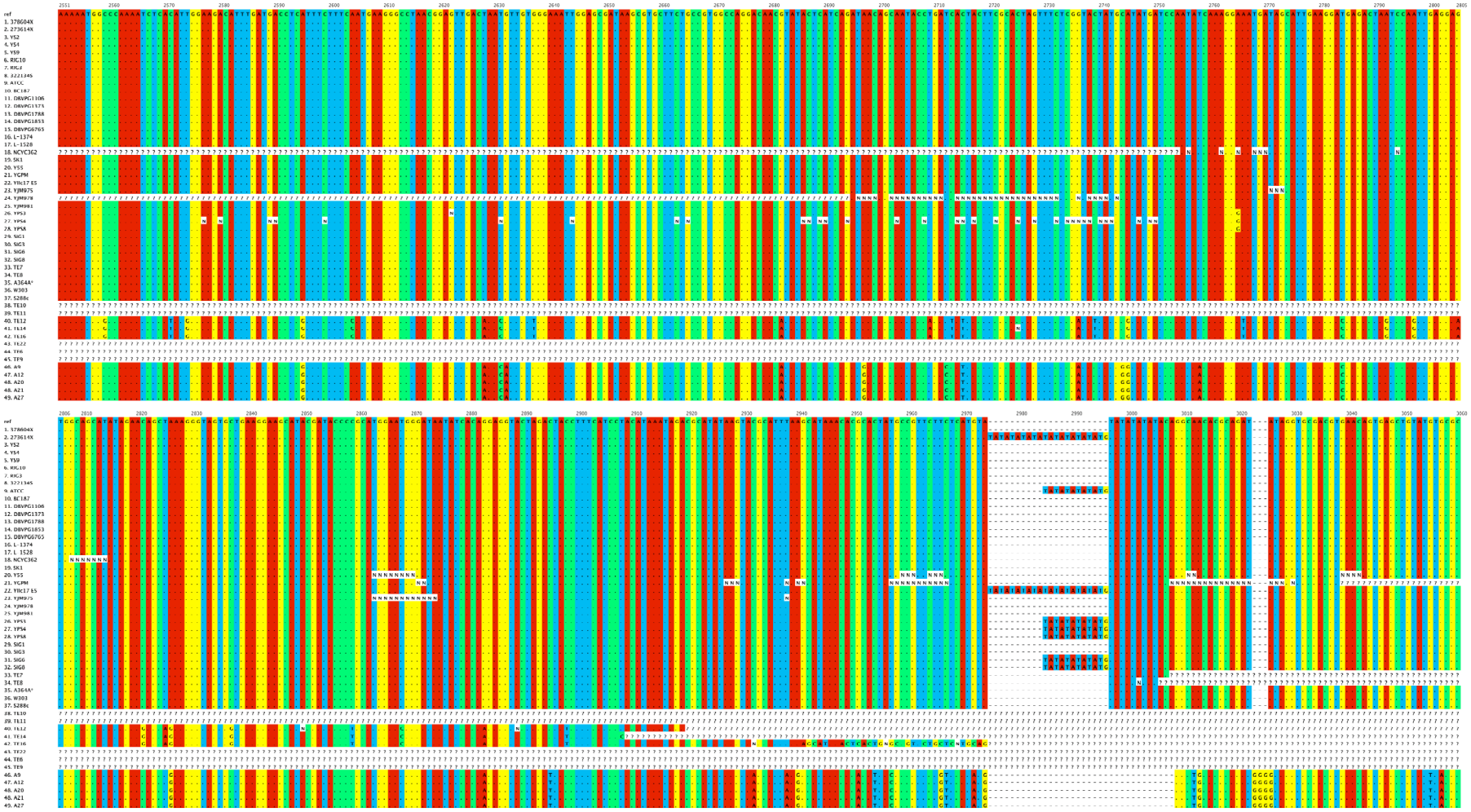


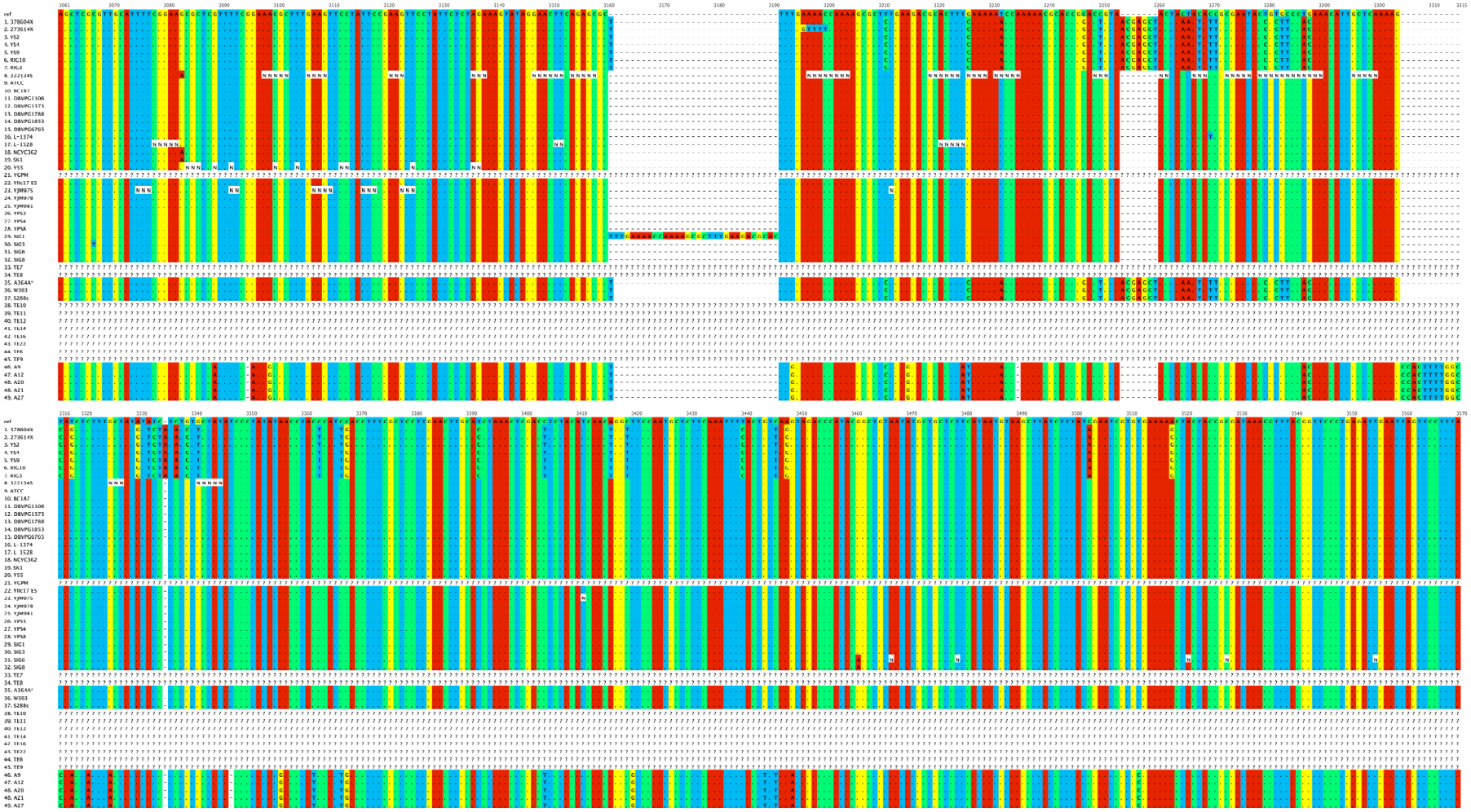
Figure 4.A.1 Complete alignment of all plasmid sequences. Sequences are named by the host strains they were isolated from and are organised into types; type 1 = 1 - 7, type 2 = 8 - 34, type 1/2 hybrids= 35 - 37, type 3= 38 - 45, type 4 = 46 – 49. Bases with quality scores (Q) <40 are given as N, all sequence gaps (–) were checked and accepted only when adjacent to bases with Q>40. ? denotes missing data due to lack of successful sequencing. Sequences are shown in comparison to a consensus sequence (Ref) and only non-matching bases are shown.

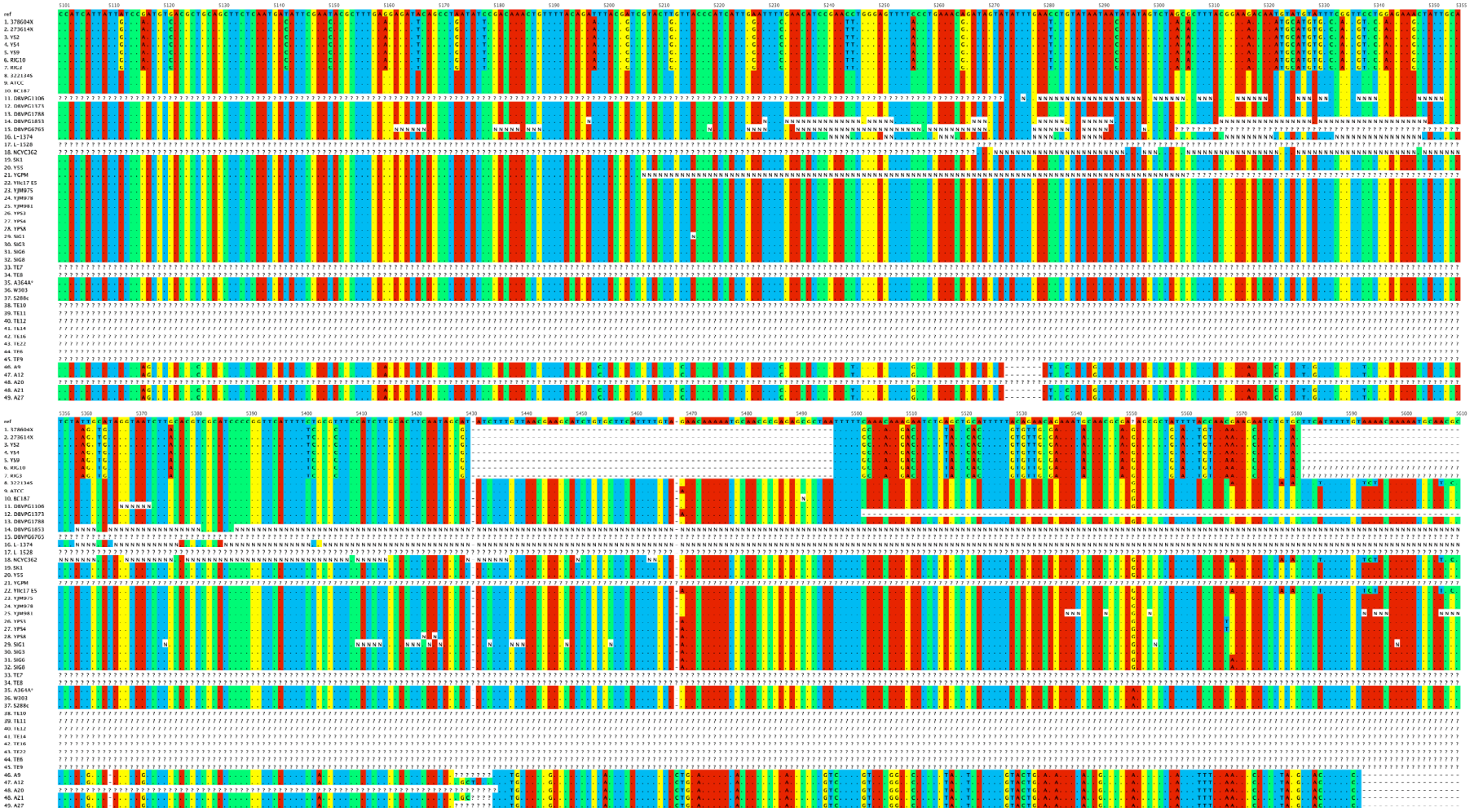


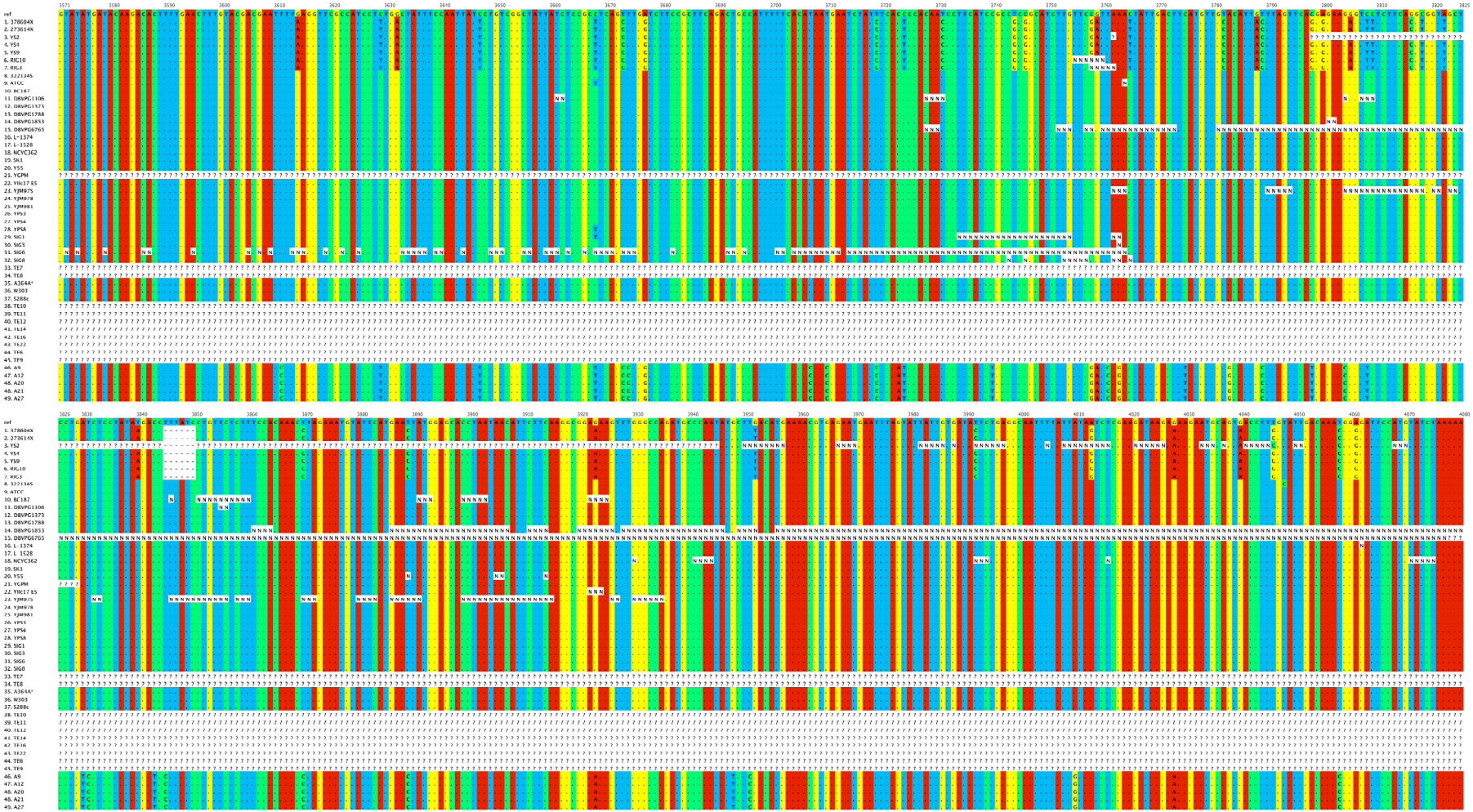


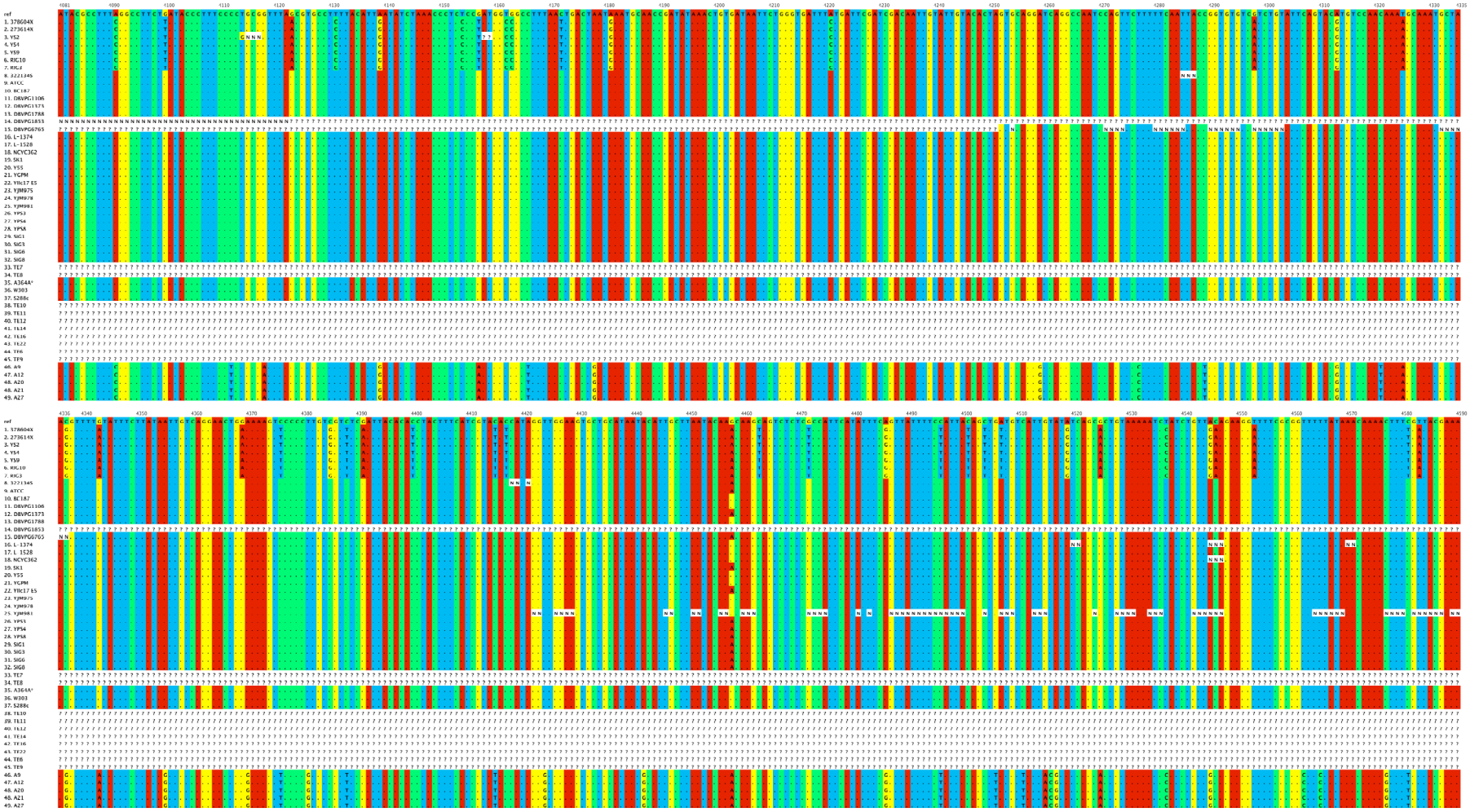


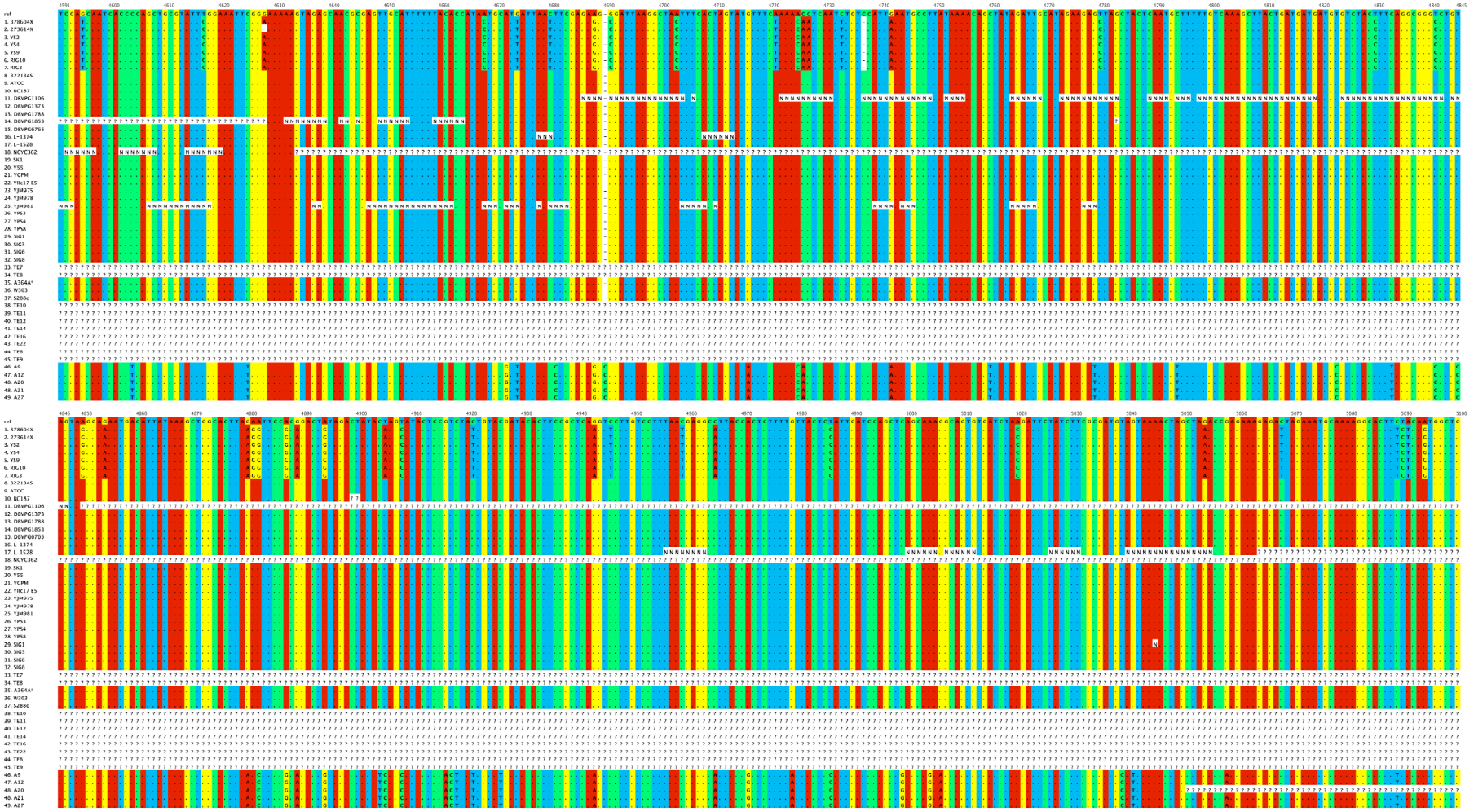


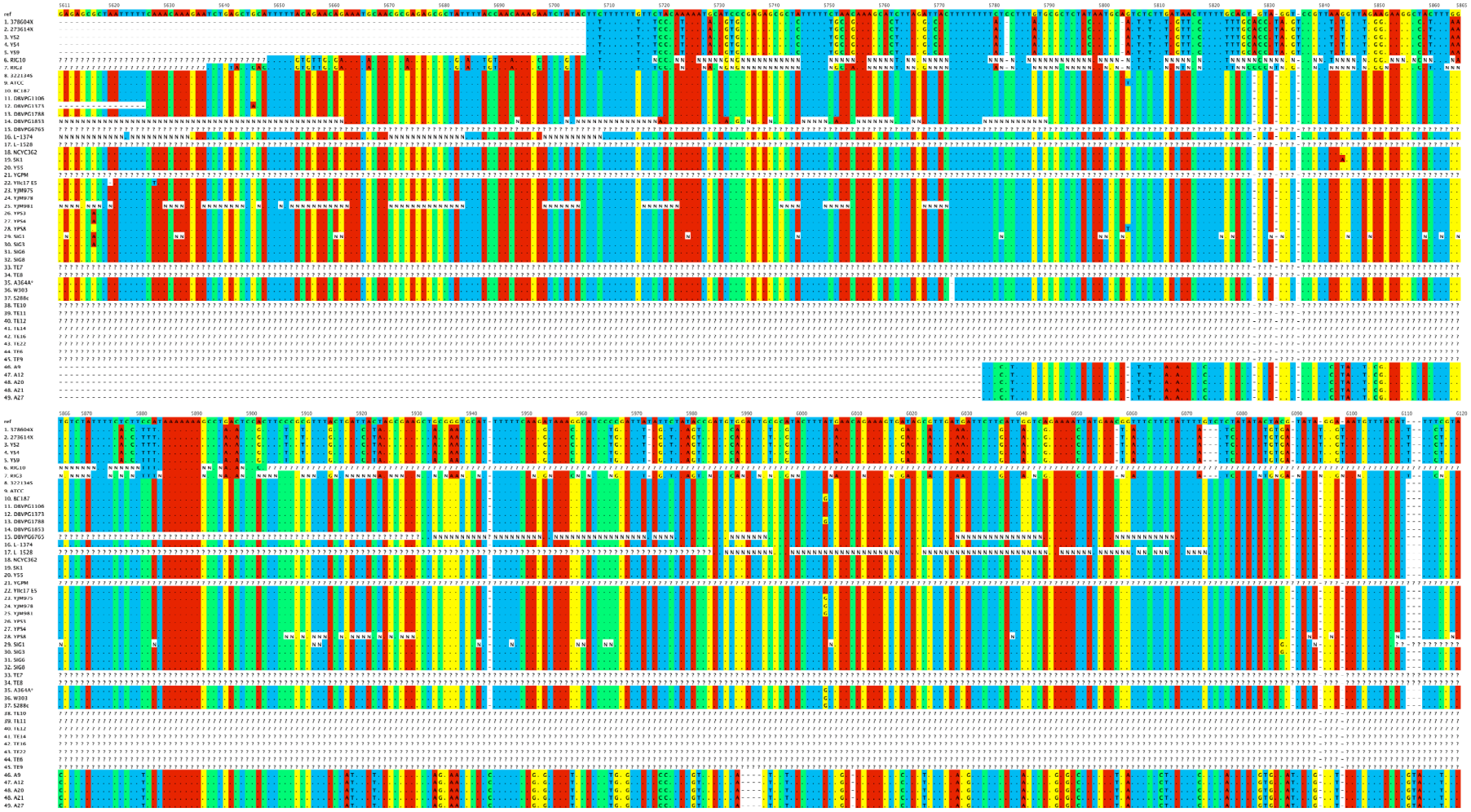


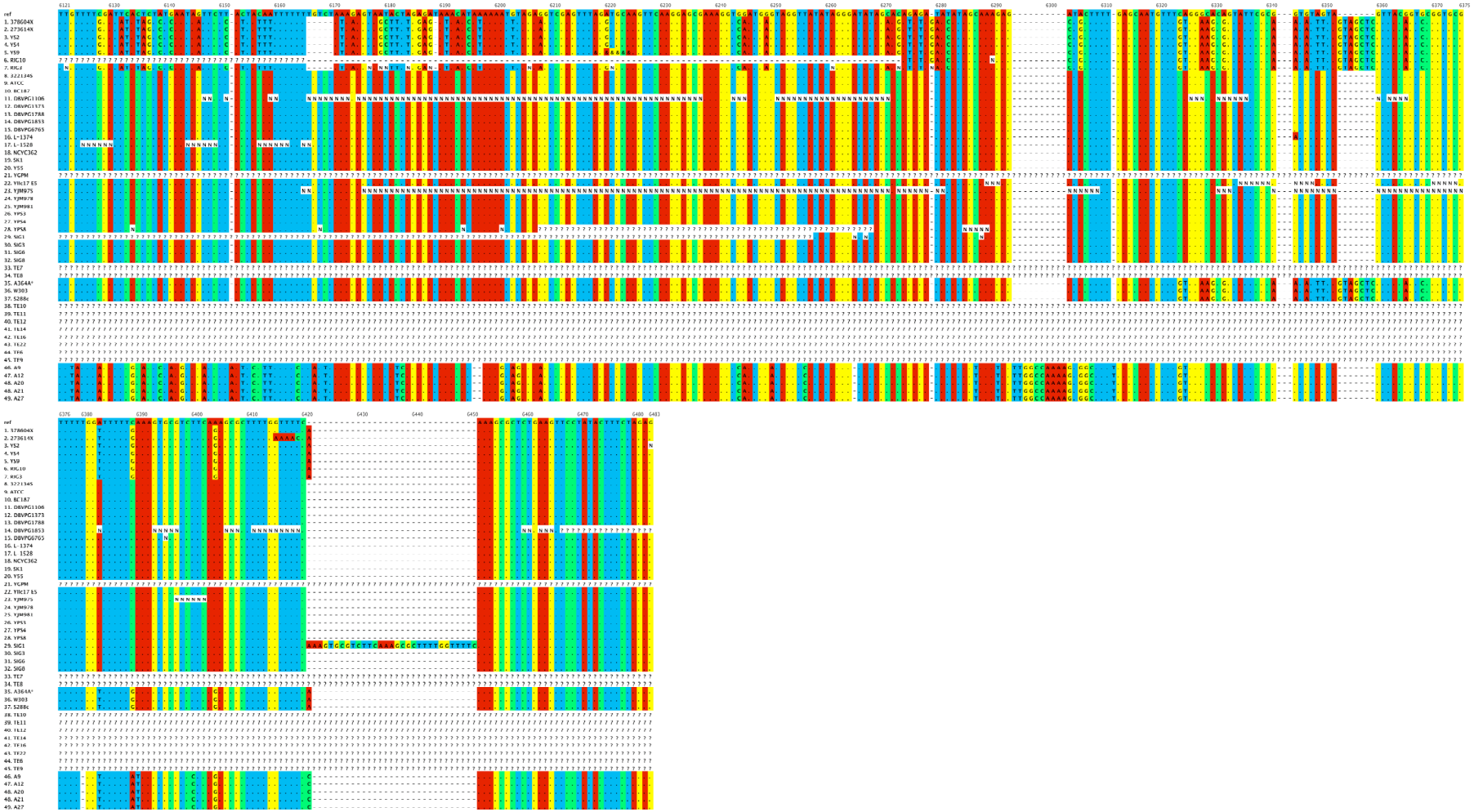












Chapter 5

Conclusions

Through the course of this thesis we have investigated the evolution of a selfish genetic element, the 2µm plasmid, which is found in the nucleus of several *Saccharomyces* yeasts (Nakayashiki et al. 2005). We have used a combination of experimental, evolutionary and comparative approaches to build a picture of the evolution and life history of this element, and to test predictions about selfish genetic elements in general.

5.1 The life and times of the 2µm plasmid

5.1.1 Copy number evolution and variation

Copy number regulation makes up one of just two mechanisms employed by the plasmid in order to ensure its stable inheritance into newly dividing cells (Velmurugan et al. 2003). However, increasing copy number incurs a fitness cost, which we have measured as ~0.09% per additional plasmid. Therefore copy number represents a measure of virulence for this element. We have demonstrated in chapter 3, that higher copy numbers evolve in populations undergoing regular outcrossed sex. This underlines that copy number evolution is driven by both, positive selection acting at the level of the gene, and negative selection acting on the individual. We proposed two sources of selection for increasing copy number. The first is that copy number effects the stability of the plasmid, as it decreases the chance of being lost through mis-segregation (Futcher & Cox 1984). However, this is likely to be a strong selective force only at low copy numbers, as in haploid cells the rate of spontaneous loss is extremely low, ~7 x 10^{-4} (Futcher et al. 1988), therefore copy number would increase with diminishing returns. A second means of selecting for increased copy number is through within host competition, as more aggressively replicating genotypes may have a selective advantage by consistently out-replicating rivals (Watve et al. 2010).

Despite the demonstrated importance of copy number to the plasmid, and the cost that it incurs, we observed a huge range of copy numbers among the 59 strains of *S. cerevisiae* and *S. paradoxus* investigated and found to contain the plasmid. Copy numbers ranged from just two up to 753 copies per diploid. These estimates are extreme considering the arguments for both positive and negative selection on copy number and therefore beg the question, how are such extremes maintained?

The lowest copy number plasmids were found in the *S. paradoxus* strains from North America. Out of 17 strains, 16 contained <4 copies, while one contained 14. If we assume that the high copy number of the plasmid is adaptive for stability, this would imply that either, the element is in the process of being lost from the population, or that low copy number plasmids are evolutionarily stable. Futcher et al. (1998) estimated that, with a rate of spontaneous loss at 7 x 10-4, and a selective advantage to [cir^0] cells at 1%, the 2μm plasmid can be maintained in a population in which out-crossing occurs only every ~5000 generations. Extremely low copy number may be sustainable in two respects. Firstly it is likely to be much less costly to the host, reducing the selective advantage of [cir^0] over [cir^+] cells and therefore, the strength of selection against the plasmid. Our estimates suggest that four plasmids reduce fitness by <0.4%. Secondly, the plasmid may potentially show greater level of stability in the host per plasmid by improved capture of cellular segregation machinery. This would mitigate the reduced stability associated with the reduction in copy number. However, analysis of the cost and stability of these plasmids would be enlightening in the face of such low numbers of this 'high copy number' plasmid.

On the contrary, the high copy number (753 copies) seen in one strain of *S. cerevisiae* inflates the total nuclear DNA by ~40%, and at our estimated rate of cost, would reduce fitness by ~68%. The appearance of such high copy number in one strain may again represent a plasmid-host association in the process of being lost from the population, for instance if the host carried a mutation in its protein modification pathways and therefore cannot effectively control copy number (Chen et al. 2005; Dobson et al. 2005; Holm 1982; Sleep et al. 2001; Xiong et al. 2009; Xu et al. 1999). However, it is also likely that selection against increasingly costly elements would result in either, greater resistance to plasmid replication and therefore a reduction in copy number, or increased tolerance to additional DNA in the cell, which would lower the cost of plasmid carriage to the host. Both these responses to plasmid burden have been observed in bacterial-plasmid associations (Lenski 1988a; Lenski 1988b; Modi & Adams 1991).

5.1.2 Genome evolution

Sequencing of the plasmid genome from 50 strains (Liti et al 2009; this study) has generated a new level of detail in the study of the 2μm genome. As had previously been shown using RFLP analysis, the type 1 and 2 plasmids are divergent at only right half of the total genome sequence, while the LUR is identical (Xiao et al. 1991; Xiao & Rank 1993; Xie et al. 1994). This pattern of divergence means that the genes and *cis*-acting regions of the RUR (RAF, REP1 and the STB locus) are all divergent between the two plasmid types, while the genes of the LUR (FLP and REP2) are conserved. Two hypotheses have followed from this finding previously. One is that divergence at the RAF, REP1 and STB region is the consequence of functional coevolution between these sites (Xiao et al. 1991). While the second is that the homology at the LUR is the consequence of polarised, directional gene conversion of this region (Xie et al. 1994).

However, sequencing of the LUR from this study has revealed small amounts of variation in the LUR, which was not observable through RFLP analysis. The phylogenetic relationships inferred using this region of the plasmid do not match that obtained using the RUR. We propose that this is evidence for recombination at the FRT region, through site-specific recombination (see figure 4.4.1), being a significant feature of the plasmids life history, analogous to sex in this small genome. In addition, we have proposed an alternative hypothesis for the observed pattern of variation between these two plasmids. We suggest that the conserved LUR may have been the product of recombination of LUR and RURs of two plasmid types, followed by a selective sweep removing one of the LURs from the population entirely (see figure 4.4.1). Therefore, we suggest that the apparent coevoluton between the RAF, REP1 and STB loci are the product of linkage, rather than adaptive coevolution at these sites.

We have also found evidence of strong selection for cheap DNA in the 2μm. DNA, and specifically, the transcription and translation of gene products are costly to the cell (Stoebel et al. 2008; Torres et al. 2007). The results from chapter 2 demonstrate that our estimate of the cost of plasmid DNA per kb is just over half that estimated for additional chromosomal DNA (Torres et al 2007). Although this estimate of cost is likely to be conservative, we also found consistently high codon bias in all four genes of the genome. This implies that the proteins encoded by he plasmid are under strong selection to be reproduced quickly and cheaply by the cell (Akashi 2003). We have argued that the cost of selfish genetic element DNA should be under strong selection to be cheap where transmission is dependent on

being frequent, as a cheap selfish genetic element can be sustained at higher frequency than an expensive rival under the same level of negative selection.

5.1.3 Population biology of the 2µm

While the molecular biology of the plasmid is extremely well understood, this represents the first study aimed at placing the plasmid in the context of its host populations, in both the wild populations of *S. paradoxus*, and in the industrial, clinical and wild strains of *S. cerevisiae*. Our results have demonstrated that the plasmid is not a benign fixture in the nucleus, but an element with its own population structure within these two species (and possibly in other members of the *Saccharomyces* genus (Nakayashiki et al. 2005)).

Four plasmid variants were identified across the two species appearing to exhibit two different kinds of population dynamics. The type 1 and 2 plasmids were isolated from strains of *S. cerevisiae* from the European/Wine clade and mosaic strains defined by Liti et al. (2009), which appear to have a history of recombination. Both plasmids are also found in European *S. paradoxus* strains from Latvia and Russia. Recombination between the two Unique Regions of these plasmids indicates that they are sympatric both in geographic location, but also occasionally in the same cell. Copy number in both type 1 and 2 plasmids is extremely variable which may be explained by the increased rate of recombination driving antagonistic intragenomic conflict between host and plasmid. In contrast, the type 3 and 4 plasmids, which have not previously been described, appear (to the extent of our survey) to be limited to their respective populations; wild Far Eastern *S. cerevisiae* and North American *S. paradoxus* strains respectively. In both populations, copy number was less variable and low. This may be due to a low frequency of outcrossing and therefore strong selection for benign elements in these populations.

5.1.4 A final word on the 2µm plasmid

The results obtained from this thesis show that the plasmid is a potentially very costly element found among the *Saccharomyces* yeast. The plasmid is extremely variable in natural populations, demonstrating its flexibility to differing conditions in the host populations. We have demonstrated that the plasmid persists in both highly recombinogenic populations in which interspecies transfer and genetic recombination between plasmids are part of the plasmid's life history, and also in reproductively isolated populations.

However the success of this plasmid in the *Saccharomyces* yeasts is not seen in any other eukaryotic species outside the budding yeasts. Both their success and their rarity is likely to lie in the use of 'point' centromeres by this group of yeasts not seen anywhere else in the eukaryotic kingdom. Malik and Henikoff (2009) have suggested that these yeasts may have obtained their distinctive point centromeres through the capture of DNA from an ancestral 2μm plasmid. If this is the case, the plasmid has been extremely fortunate as it is the mimicry of the centromeric system that has allowed the plasmid to be so successful in this small family of yeasts.

5.2 Sex and the evolution of Selfish genetic elements

This thesis has also sought to address broader questions about the evolution of selfish genetic elements in general. The evolution of selfish genetic elements, which use solely vertical modes of transmission, is inextricably linked to sex (discussed in Chapters 1 and 3). We have demonstrated, through a long term selection experiment, that sex can drive the evolution of selfish genetic elements, both in terms of their emergence and spread in the case of the Selfish Mitochondria, and in evolving increasing virulence in the 2μm plasmid. To our knowledge, this is the first experimental demonstration of how host sexual system drives increasing levels of virulence in selfish genetic elements. The impacts of these finding on selfish genetic element evolution have been discussed in chapter 3, however they may also be relevant to a number of other areas of evolutionary biology which I will discuss here.

5.2.1 Levels of Selection

The results presented in Chapter 3 provide an example of a conflict between two levels of selection driving evolution. In asexual organisms the genome is the most basic level of selection, as it is inherited as a single unit. For sexual organisms, recombination breaks up the genome into linked groups of genes therefore, giving rise to within-individual selection between these genes (Hurst 1992). Selfish genetic elements represent genetic units which have gained a transmission advantage over the rest of the genome, called drive (Burt & Trivers 2006). Drive weakens the strength of between-individual selection, as elements are able to increase in frequency in the population without relying on positive selection as is the case for most of the genome (Hickey 1982). In addition, sex decreases the relatedness within the genome increasing the strength of within-individual selection through competition between elements.

The effect of conflicting levels of selection on the evolution of Selfish Mitochondria has been demonstrated in this system previously by Taylor et al (2002). Selfish Mitochondria were unable to invade large populations, in which between-individual selection is effective. However, in nearly all small populations, Selfish Mitochondria went to fixation, as stochastic processes such as genetic drift, rather natural selection dominate the changes in gene frequencies. Our results support the conclusions of Taylor et al (2002), that the evolution of Selfish Mitochondria, and selfish genetic elements in general are driven by conflicts between these levels of selection.

5.2.2 Evolution of Endosymbiosis

The evolution of endosymbionts is a process plagued with the potential for conflict. In sexually reproducing organisms the fusion of gametes means that cytoplasmic elements with a replication advantage are able to spread in the population in the face of negative selection (i.e. drive) (Hastings 1992). However, this conflict is prevented by enforcing uniparental inheritance in cytoplasmic elements (Hurst 1995), which prevents the spread of selfish variants. But this comes at a cost, as the evolution of sexes (in which one sacrifices its cytoplasm) results in the "two-fold cost of sex" (Hastings 1992). The spread of Selfish Mitochondria into sexual populations in chapter 3 supports the hypothesis that sex drives the evolution of costly cytoplasmic elements when cytoplasm is bi-parentally inherited. Therefore we can assume that organelles like mitochondria and chloroplasts evolved under asexual conditions. As sex evolved and increased in frequency, increasing conflict between cytoplasmic elements could support the spread of alleles favouring the division of the sexes (Hastings 1992).

5.2.3 Evolution of Sex

The evolution of sexual reproduction in eukaryotes remains one of the presiding mysteries in evolutionary biology (Charlesworth 2006). We have demonstrated that regular sexual reproduction can drive the spread of selfish genetic elements, and their evolution of higher virulence. In these populations, sexual reproduction did not result in faster adaptation relative to asexual populations, as has been shown in a similar experiment (Goddard et al. 2005), suggesting that perhaps, the rising levels of conflict within the genome is sufficient to negate the adaptive benefits of sex. This can also be seen by the increased prevalence of selfish genetic elements in sexual isolates of Daphnia pulex compared to obligate asexuals

(Schaack et al. 2010; Valizadeh & Crease 2008), and in the majority of rotifer clades compared to the asexual Bdelloid rotifers (Arkhipova & Meselson 2000; Arkhipova 2005).

This increased burden can be costly in two ways, both as a direct result of increased selfish genetic element activity, and through selection for resistance which can itself incur a cost (Boots & Begon 1993; Kraaijeveld et al. 2002). The evolution of resistance to selfish genetic elements may have driven theevolution of gene silencing mechanisms, RNAi and methylation (Johnson 2007) as well as the evolution of uniparental inheritance (Hurst 1995). Therefore, selfish genetic elements have clearly been important in shaping the evolution of eukaryotes as sexuality increased.

5.3 Conclusions

The evolution of selfish genetic elements is a fascinating and complex subject, which is being increasingly found to play a major role in eukaryotic evolution in general. A great deal is yet to be learned about the life histories, biology and evolution of these elements, making their study hugely rewarding. The work presented here has focused on the evolution of just one element, the 2μm plasmid, which, in its microbial host, provides a powerful model system for testing broader evolutionary questions. Here we have attempted to address both questions specific to this system, but also test predictions applicable to the evolution of selfish genetic elements in general.

6. Appendix I

General Materials and Methods

6.1 Strains

Table A1 Experimental *S. cerevisiae* strains used in chapters 1 and 2

Name	Genotype	Reference
YDG 598	*a/α ho/ho his/his leu/leu ura/ura SUC2/SUC2*	
YDG 599	*a/α* ho/ho his/his leu/leu ura/ura SUC2*Δ*::KanMX/SUC2*Δ*::KanMX	
MG 98 (A)	*a/α ho/ho ura3Δ/ura3Δ spo11Δ/spo11Δ spo13Δ::KanMX4/spo13Δ::KanMX4*	
MG 99	*a/α ho/ho ura3Δ/ura3Δ*	(Goddard et al. 2005)
DH 89e (S)	*ho/ho a/α URA3/ura3Δ lys2Δ/LYS2*	

Table A2 Strains used to survey for the presence of the 2 micron plasmid.

Species	Strain	Type	Origin	Sourse
S. cerevisiae	273614X	Clinical	Newcastle UK	V. Koufopanou
	322134S	Clinical	Newcastle UK	V. Koufopanou
	378604X	Clinical	Newcastle UK	V. Koufopanou
	BC187	Fermentation	Napa Valley, USA	V. Koufopanou
	BCRC 20270	Clinical		V. Koufopanou
	DBVPG1106	Fermentation	Australia	V. Koufopanou
	DBVPG1373	Wild	Netherlands	V. Koufopanou
	DBVPG1788*	Wild	Finland	V. Koufopanou
	DBVPG1853	Wild	Ethiopia	V. Koufopanou
	DBVPG6040*	Fermentation	Netherlands	V. Koufopanou
	DBVPG6044	Fermentation	West Africa	V. Koufopanou
	DBVPG6765	Unknown	Unknown	V. Koufopanou
	K11	Fermentation	Japan	V. Koufopanou
	L1374	Fermentation	Chile	V. Koufopanou
	L1528	Fermentation	Chile	V. Koufopanou
	NCYC110	Fermentation	West Africa	V. Koufopanou
	NCYC361	Fermentation	Ireland	V. Koufopanou
	S288c	Wild	California, USA	V. Koufopanou
	SK1*	Wild	USA	V. Koufopanou
	TE6 (TCJ 11)	Wild	Taiwan	C.R Tien
	TE7 (TCJ 21)	Wild	Taiwan	C.R Tien
	TE8 (TCJ 22)	Wild	Taiwan	C.R Tien
	TE9 (TCJ 31)	Wild	Taiwan	C.R Tien
	TE10 (TCJ 85)	Wild	Taiwan	C.R Tien

	TE11 (TCJ 87)	Wild	Taiwan	C.R Tien
	TE12 (TCJ 90)	Wild	Taiwan	C.R Tien
	TE12a (TCJ 92)	Wild	Taiwan	C.R Tien
	TE13 (TCJ 126)	Wild	China	C.R Tien
	TE14 (TCJ 152)	Wild	China	C.R Tien
	TE15 (TCJ 153)	Wild	China	C.R Tien
	TE16 (TCJ 154)	Wild	China	C.R Tien
	TE17 (TU-6-1)	Wild	Taiwan	C.R Tien
	TE18 (TU-8-1)	Wild	Taiwan	C.R Tien
	TE19 (TU-9-1)	Wild	Taiwan	C.R Tien
	TE21 (012409-18)	Wild	Taiwan	C.R Tien
	TE22 (012409-19)	Wild	Taiwan	C.R Tien
	UWOPS03-461.4	Wild	Malaysia	V. Koufopanou
	UWOPS05-217.3	Wild	Malaysia	V. Koufopanou
	UWOPS05-227.2	Wild	Malaysia	V. Koufopanou
	UWOPS83-787.3	Wild	Bahamas	V. Koufopanou
	UWOPS87-2421	Wild	Hawaii	V. Koufopanou
	W303	Unknown	Unknown	V. Koufopanou
	Y12	Fermentation	Africa	V. Koufopanou
	Y55	Fermentation	France	V. Koufopanou
	Y9		Japan	V. Koufopanou
	YIIc17	Fermentation	Sauternes, France	V. Koufopanou
	YJM975	Clinical	Bergamo, Italy	V. Koufopanou
	YJM978	Clinical	Bergamo, Italy	V. Koufopanou
	YJM981	Clinical	Bergamo, Italy	V. Koufopanou
	YS2	Baking	Australia	V. Koufopanou
	YS9	Fermentation	Singapore	V. Koufopanou
	YS4	Baking	Netherland	V. Koufopanou
	YPS 128	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 128	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 129	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 133	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 134	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 139	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 142	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 143	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 143	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 163	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 606	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 627	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 629	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 670	Wild	Pennsylvania, USA	P. Sniegowski
	YPS 681	Wild	Pennsylvania, USA	P. Sniegowski
S. paradoxus	CECT10308	Wild	J. Santa Maria , Spain	V. Koufopanou
	CECT10329*	Wild	J. Santa Maria , Spain	V. Koufopanou
	SIG 1	Wild	Sigulda, Latvia	V. Koufopanou
	SIG 8	Wild	Sigulda, Latvia	V. Koufopanou
	ATCC 96975	Wild	Moscow, Russia	V. Koufopanou
	ATCC 96983	Wild	Moscow, Russia	V. Koufopanou

ATCC 96986	Wild	Moscow, Russia	V. Koufopanou
CBS 8436	Wild	Vladivostok, Russia,	V. Koufopanou
CBS 8437*	Wild	Vladivostok, Russia,	V. Koufopanou
CBS 8438*	Wild	Ternei City, Russia	V. Koufopanou
CBS 8439*	Wild	Ternei City, Russia	V. Koufopanou
CBS 8440	Wild	Sikote-Alinsky nature reserve, Russia	V. Koufopanou
CBS 8441	Wild	Sikote-Alinsky nature reserve, Russia	V. Koufopanou
CBS 8442	Wild	Khazansky District, Russia	V. Koufopanou
CBS 8443	Wild	Khazansky District, Russia	V. Koufopanou
CBS 8444	Wild	Vladivostok, Russia,	V. Koufopanou
CBS5829	Wild	Denmark	V. Koufopanou
CECT10176	Wild	J. Santa Maria , Spain	V. Koufopanou
CECT10178*	Wild	J. Santa Maria , Spain	V. Koufopanou
DBVPG4650	Wild	Italy	V. Koufopanou
IFO1804	Wild	Japan	V. Koufopanou
KPN3829	Wild	Novosibirsk, Siberia, Russia,	V. Koufopanou
N17*	Wild	Russia	V. Koufopanou
YPS 12	Wild		V. Koufopanou
YPS 13	Wild		V. Koufopanou
YPS 3	Wild	Leningrad, Russia	V. Koufopanou
YPS 4	Wild	Moscow, Russia	V. Koufopanou
YPS 6	Wild		V. Koufopanou
YPS 7	Wild		V. Koufopanou
YPS 8	Wild	Moscow, Russia	V. Koufopanou
YPS 9	Wild		V. Koufopanou
A20 20 Soil S	Wild	St. Halaire, Canada	G. Bell
A36 36 Soil S	Wild	St. Halaire, Canada	G. Bell
A7 7 Soil W	Wild	St. Halaire, Canada	G. Bell
J10 10 SLR	Wild	St. Halaire, Canada	G. Bell
A11 11 1S	Wild	St. Halaire, Canada	G. Bell
A13 13 2W	Wild	St. Halaire, Canada	G. Bell
A15 15 Soil W	Wild	St. Halaire, Canada	G. Bell
A17 17 Soil S	Wild	St. Halaire, Canada	G. Bell
A19 19 2S	Wild	St. Halaire, Canada	G. Bell
A28 28 2W	Wild	St. Halaire, Canada	G. Bell
A29 29 2W	Wild	St. Halaire, Canada	G. Bell
A30 30 2W	Wild	St. Halaire, Canada	G. Bell
A31 31 1W	Wild	St. Halaire, Canada	G. Bell
A32 32 1S	Wild	St. Halaire, Canada	G. Bell
J10 10 ELR	Wild	St. Halaire, Canada	G. Bell
J2 2 Soil S1	Wild	St. Halaire, Canada	G. Bell
J7 7 Soil W2	Wild	St. Halaire, Canada	G. Bell
J9 9 NLT	Wild	St. Halaire, Canada	G. Bell
J9 9 WLT	Wild	St. Halaire, Canada	G. Bell
J3 3 Soil W2	Wild	St. Halaire, Canada	G. Bell
J3 3 SLT	Wild	St. Halaire, Canada	G. Bell
J4 4 NLR	Wild	St. Halaire, Canada	G. Bell
J4 4 Soil E2	Wild	St. Halaire, Canada	G. Bell
J5 5 Soil E2	Wild	St. Halaire, Canada	G. Bell

J5 5 Soil W1	Wild	St. Halaire, Canada	G. Bell
J8 8 SLT	Wild	St. Halaire, Canada	G. Bell
Q14.4	Wild	Silwood Park, UK	V. Koufopanou
Q15.1	Wild	Silwood Park, UK	V. Koufopanou
Q16.1	Wild	Silwood Park, UK	V. Koufopanou
Q31.4	Wild	Silwood Park, UK	V. Koufopanou
Q32.3	Wild	Silwood Park, UK	V. Koufopanou
Q4.1	Wild	Silwood Park, UK	V. Koufopanou
Q43.5	Wild	Silwood Park, UK	V. Koufopanou
Q59.1	Wild	Silwood Park, UK	V. Koufopanou
Q6.1	Wild	Silwood Park, UK	V. Koufopanou
Q62.5	Wild	Silwood Park, UK	V. Koufopanou
Q69.8	Wild	Silwood Park, UK	V. Koufopanou
Q70.8	Wild	Silwood Park, UK	V. Koufopanou
Q74.4	Wild	Silwood Park, UK	V. Koufopanou
Q89.8	Wild	Silwood Park, UK	V. Koufopanou
Q95.3	Wild	Silwood Park, UK	V. Koufopanou
RIG 10	Wild	Riga, Latvia	V. Koufopanou
RIG 11	Wild	Riga, Latvia	V. Koufopanou
RIG 3	Wild	Riga, Latvia	V. Koufopanou
S36.7*	Wild	Silwood Park, UK	V. Koufopanou
SIG 3	Wild	Sigulda, Latvia	V. Koufopanou
SIG 6	Wild	Sigulda, Latvia	V. Koufopanou
STOC 3	Wild	Stockholm, Sweden	V. Koufopanou
T18.2	Wild	Silwood Park, UK	V. Koufopanou
T21.4	Wild	Silwood Park, UK	V. Koufopanou
T26.3	Wild	Silwood Park, UK	V. Koufopanou
T46	Wild	Silwood Park, UK	V. Koufopanou
T62.1	Wild	Silwood Park, UK	V. Koufopanou
T68.1	Wild	Silwood Park, UK	V. Koufopanou
T68.2	Wild	Silwood Park, UK	V. Koufopanou
T76.6	Wild	Silwood Park, UK	V. Koufopanou
T8.1	Wild	Silwood Park, UK	V. Koufopanou
W7*	Wild	Silwood Park, UK	V. Koufopanou
Y1	Wild	Silwood Park, UK	V. Koufopanou
Y10	Wild	Silwood Park, UK	V. Koufopanou
Y2	Wild	Silwood Park, UK	V. Koufopanou
Y2.2	Wild	Silwood Park, UK	V. Koufopanou
Y2.8	Wild	Silwood Park, UK	V. Koufopanou
Y3	Wild	Silwood Park, UK	V. Koufopanou
Y4	Wild	Silwood Park, UK	V. Koufopanou
Y4.5	Wild	Silwood Park, UK	V. Koufopanou
Y5	Wild	Silwood Park, UK	V. Koufopanou
Y5.1	Wild	Silwood Park, UK	V. Koufopanou
Y5.6	Wild	Silwood Park, UK	V. Koufopanou
Y6	Wild	Silwood Park, UK	V. Koufopanou
Y6.5	Wild	Silwood Park, UK	V. Koufopanou
Y7	Wild	Silwood Park, UK	V. Koufopanou
Y7.1	Wild	Silwood Park, UK	V. Koufopanou

Y8.1	Wild	Silwood Park, UK	V. Koufopanou
Y8.4	Wild	Silwood Park, UK	V. Koufopanou
Y8.5	Wild	Silwood Park, UK	V. Koufopanou
Y8.6	Wild	Silwood Park, UK	V. Koufopanou
Y8.8	Wild	Silwood Park, UK	V. Koufopanou
Y9	Wild	Silwood Park, UK	V. Koufopanou
Y9.5	Wild	Silwood Park, UK	V. Koufopanou
Z1	Wild	Silwood Park, UK	V. Koufopanou
Z1.1	Wild	Silwood Park, UK	V. Koufopanou

6.2. Microbiology

Table A3 List of commonly used medias

YPD :	10g/L	Yeast Extract
	20g/L	Peptone
	20g/L	Glucose
	(16g/L	Agar)
YPG :	10g/L	Yeast Extract
	20g/L	Peptone
	20g/L	Glycerol
	(16g/L	Agar)
Minimal Media (MM) :	1.7g/L	Yeast Nitrogen Base (without amino acids or Ammonium Sulfate)
	5g/L	Ammonium Sulfate
	20g/L	Glucose
	(16g/L	Agar)
Synthetic complete (SC) :	1.4g/L	Minimal media plus Dropout mix (Sigma)
LB :	5g/L	Yeast Extract
	10g/L	Peptone
	10g/L	NaCl
	(16g/L	Agar)
Additions	2g/L	αAminoadipate (use in MM without ammonium sulphate)
	1g/L	5- Fluoroorotic Acid
	0.02g/L	Uracil
	0.02g/L	Histadine
	0.05g/L	Tryptophan
	0.1g/L	Leucine
	0.1g/L	Lysine
Sporulation media :	5g/L	Potassium Acetate
Outcrossing solution :	80 U/ml	Lyticase
	50mM	Dithiotritol
	1100 U/ml	β-Glucoronidase
Saline :	0.9g/L	Sodium Chloride (NaCl)

6.3 Molecular biology

Table A34 List of commonly used reagents

Lysis Buffer :	1%	SDS
	50mM	EDTA
	0.1M	Tris-HCl
TE :	10mM	Tris-HCl
	1mM	EDTA
	pH8	
Lyticase Solution :	1M	Sorbitol
	1mM	EDTA
	10mM	Tris-HCl
	40 U/ml	Lyticase (filter sterilized and added to solution post autoclaving)
Transformation solution :	2%	Tritol-X
	1%	SDS
	100mM	NaCl
	10mM	Tris- HCl (pH 8.0)
	1mM	EDTA

DNA Extractions from yeast

For 1-2ml individual cultures

Based on the Promega Wizard Genomic DNA purification Kit

- Centrifuge (13k*g,* 1min) and remove supernatant.
- Resuspend in 0.5ml Diploid digestion solution and incubate at 37^{o}C for 2hrs.
- Centrifuge (13k*g,* 1min), remove supernatant and add 400μl Lysis buffer and a 600 micron acid washed glass beads.
- Vortex for 5mins
- Add 300μl of suernatant to 100μl Protein precipitate solution (Promega)
- Centrifuge (13k*g,* 3mins), transfer 300μl supernatant to a new centrifuge tube containing 300μl 100% Ethanol.
- Mix by inversion and incubate on ice for >20mins.
- Centrifuge (13k*g,* 2mins) to precipitate DNA.
- Pour off supernatant, add 0.75ml 70% ethanol and mix by inversion.
- Centrifuge (13k*g,* 2mins), pour off ethanol and leave at 37^{o}C to dry overnight.

- Add 50µl TE or ddH_2O and 1.5µl RNase (Promega), incubate at 37°C for 15mins.
- Incubate at 65°C for 1hr or at 4°C overnight.

Southern Blotting

Blotting was done using the DIG High Prime DNA Labeling and Detection Starter Kit I (Roche). Probes were made from PCR products of primer sets REP1, REP2, FLP, RAF, STB and the genomic control ITS as described below. Products were cleaned using Promega Wizard SV Gel/PCR Clean Up and labeled as described in the protocol provided and optimized according to recommendations in the manual.

1µg restriction digested DNA was run on a gel made up of a 2% agarose base (for stability) and a 0.6% layer through which DNA is allowed to run. Gels were run for >7hrs at 20V.. A Hybond N+ membrane (Roche) was cut to fit the gel and the well positions marked on the membrane.

The gel was washed in H_2O and then depurinated in 0.25M HCl for 20mins under constant agitation. It was then neutralized in 0.4M NaOH for 30mins and the Hybond N+ membrane incubated in 0.4M NaOH for 10mins.

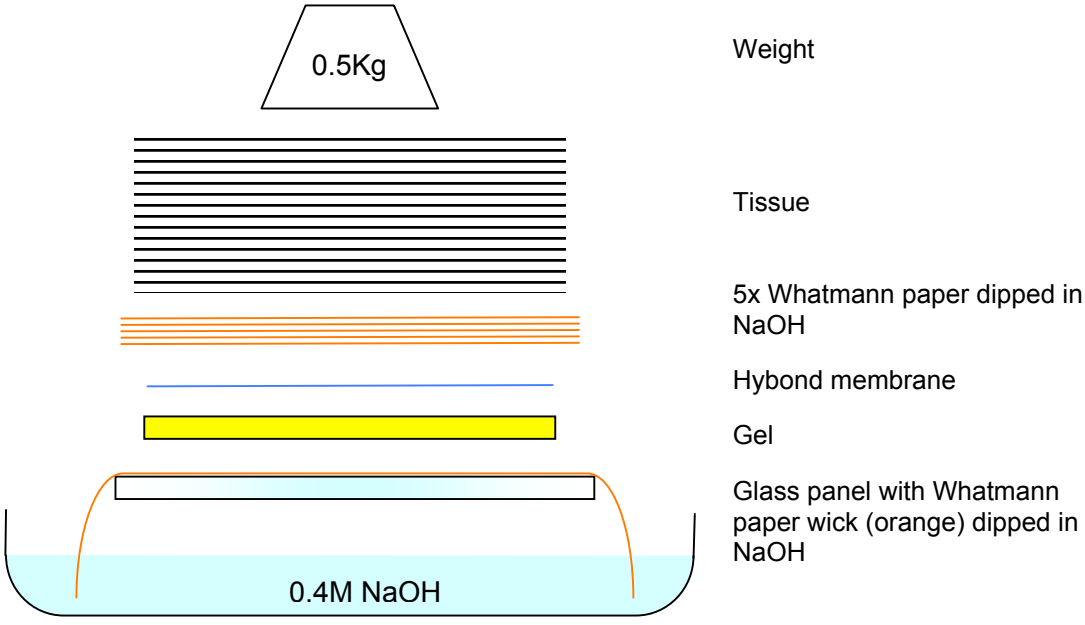


Figure A1 Construction of southern blotting apparatus

DNA was transferred overnight by capillary action using the method shown in figure A.1.

Whatmann paper was dipped in NaOH solution and layed over the glass base. Any bubbles were removed with a glass rod. The gel was placed DNA side up and clingfilm used to block any transfer of liquid around (rather than through) the gel. The rest of the stack was layed on top as seen above, making sure no air bubbles were present between wet layers.

Following over night transfer of DNA the membrane was recovered and rinsed in 2x SSC (17.53g/L NaCl, 8.82g/L trisodium citrate) and left to dry between 2 sheets of Whatmann paper overnight.

Hybridisation and visualization of the bands proceeded as described in the kit protocol using a hybridization temperature of 37^{o}C for 4 hrs. For plasmid detection the probe was constructed from 20µl of each of the 2µm primer probes- therefore a total of 100µl of probe was added. For the control blot 50µl of the ITS probe was added.

PCR

Table A4 Commonly used PCR primers including reaction conditions

Name		Sequence	Annealing temp	Extension time	Cycle no.
FLP	for	TGCTTCCTTCAGCACTACC	57	2	40
	rev	CACTACTTCGCACTAGTTTCTC			
FLP_3	for	ACAGCCGACATACATCCCAG	57	2	30
	rev	GCGCTTCCGAAAATGCAACG			
RAF	for	ACGATACACTTCCGCTCAG	55	2	40
	rev	TGCCTTTTGCATTTCTAGTCTC			
REP1	for	TCACCCCACAATCCTTCATC	55	2	40
	rev	CATCTGGCCCAAACTTCTC			
REP2_2	for	GCTGGTGGGACTAATAACTGTG	57	2	40
	rev	ACGACATTGAAACAGCCAAG			
STB_2	for	TGCGTTTCCATCTTGCAC	55	2	40
	rev	GAACAGAAATGCAACGCG			
ITS	1	TCCGTAGGTGAACCTGCGG	55	2	40
	4	TCCTCCGCTTATTGATATGC			

Standard PCR:

Reactions were performed with either BioRed (Bioline) or Accuzyme (Bioline) reaction mix in a Biometra Tgradient thermocycler.

For a 25μl reaction:

BioRed/Accuzyme	12.5μl
Forward primer (10pmol)	0.5μl
Reverse primer (10pmol)	0.5μl
H_2O	10.5μl
Template	1μl

95°C	2mins	
95°C	30s	
X°C	1min	x cycle number
72°C	Xmins	
72°C	5mins	
4°C	pause	

Colony PCR:

Yeast colonies were grown on appropriate solid media for 2 days. >10μl cells were taken from each colony and resuspended in 20μl ddH_2O in 50ul PCR tubes/96 well plate. Pick yeast from plate (I use a toothpickful) and suspend in 20ul dH_20. These were placed in the thermocycler for 10mins at 100^0C to lyse cells. Tubes were then spun down at 13k*g* for 1 min, placed on ice and 1μl used as template immediately in the standard PCR mix described above. Reaction cycles was the same as for standard PCR but with an initial time of 15mins at 95°C

Quantitative PCR:

All primers were designed using Primerexpress software, which picks primers optimal for the ABIprism qPCR programme. All qPCRs were run in the ABI sequence7000 PCR machine. SensiMix Real Time PCR reaction mix (Bioline) was used following manufactures reaction conditions. Primer concentrations were optimised but in all cases 300pmol concentrations provided the best results.

Calculations

Copy number is calculated from the Ct (threshold cycle) of the amplification curves, being the cycle at which a samples florescence passes a given threshold (1.0 used in this case).

Each plate includes a dilution series of at least five ten-fold dilutions. Standard curves of Ct by DNA concentration were then constructed and used to calibrate all calculations of copy number. The slope of the standard curve represents the efficiency of the PCR reaction and is optimal at –3.33. Slopes greater than –3 and lower than –3.66 were adjusted by removing the lowest or highest DNA concentrations and then rejected if still not within these limits. Individual Ct values greater or less than those included in the standard curve were not included in the analysis.

Absolute copy number was calculated for the standard strain using the dilution series on each plate.

(1) Ct plotted against log[DNA(ng/reaction)] for plasmid (P) and genomic (G) primers

(2) Absolute copy number per diploid= $(10^{(\text{INTERCEPT(P)} - \text{INTERCEPT(G)})/\text{SLOPE}}) \times 2$

Individual samples were calculated as relative copy numbers, using $\Delta\Delta$Ct method (Lee et al. 2006), as it requires less reactions per sample. The relative method uses a comparison between the sample and a standard. For this the 10ng/reaction dilution of Y55 standard curve was used.

(3) $\Delta Ct = Ct(P) - Ct(G)$

(4) $\Delta\Delta Ct = \Delta Ct(\text{SAMPLE}) - \Delta Ct(\text{STANDARD})$

(5) Relative copy number = $(1 + E)^{-\Delta\Delta Ct}$

E is the amplification efficiency as calculated from the standard curve

(6) $E = 10^{-1/\text{SLOPE}} - 1$

Copy number estimates are then calculated as the relative value times the absolute value of the standard strain and averages given are the result of this estimate for each replicate.

7. References

Agrawal, A., Eastman, Q. M. & Schatz, D. G. 1998 Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394, 744-751.

Akashi, H. 1994 Synonymous codon usage in Drosophila melanogaster - Natural selection and translational accuracy *Genetics* 136, 927-935.

Akashi, H. 2003 Translational selection and yeast proteome evolution. *Genetics* 164, 1291-1303.

Alfieri, J. A. & Clark, D. J. 1999 Isolation of minichromosomes from yeast cells. In *Chromatin*, vol. 304, pp. 35-49.

Amberg, D. C., Burke, D. & Strathern, J. N. 2005 *Methods in Yeast Genetics*: Cold Spring Harbor Laboratory Press.

Araki, H., Jearnpipatkul, A., Tatsumi, H., Sakurai, T., Ushio, K., Muta, T. & Oshima, Y. 1985 Molecular and Functional Organization of Yeast Plasmid pSR1. *Journal of Molecular Biology* 182, 191-203.

Archetti, M. 2003 A selfish origin for recombination. *Journal of Theoretical Biology* 223, 335-346.

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

Arkhipova, I. R. 2005 Mobile genetic elements and sexual reproduction. *Cytogenetic and Genome Research* 110, 372-382.

Bensasson, D., Zarowiecki, M., Burt, A. & Koufopanou, V. 2008 Rapid evolution of yeast centromeres in the absence of drive. *Genetics* 178, 2161-2167.

Bestor, T. H. 1999 Sex brings transposons and genomes into conflict. *Genetica* 107, 289-295.

Bilinski, C. A., Russell, I. & Stewart, G. G. 1986 Analysis of Sporulation in Brewers-Yeast - Induction of Tetrad Formation. *Journal of the Institute of Brewing* 92, 594-598.

Blaisonneau, J., Sor, F., Cheret, G., Yarrow, D. & Fukuhara, H. 1997 A circular plasmid from the yeast Torulaspora delbrueckii. *Plasmid* 38, 202-209.

Boots, M. & Begon, M. 1993 Trade-offs with resistance to a granulosis-virus in the Indian meal moth, examined by a laboratory evolution experiment. *Functional Ecology* 7, 528-534.

Bouma, J. E. & Lenski, R. E. 1988 Evolution of a bacteria plasmid association. *Nature* 335, 351-352.

Buckling, A. & Brockhurst, M. A. 2008 Kin selection and the evolution of virulence. *Heredity* 100, 484-488.

Buckling, A. & Rainey, P. B. 2002 Antagonistic coevolution between a bacterium and a bacteriophage. *Proceedings of the Royal Society of London Series B-Biological Sciences* 269, 931-936.

Bull, J. J. 1994 Perspective - Virulence. *Evolution* 48, 1423-1437.

Burgess, R. C., Rahman, S., Lisby, M., Rothstein, R. & Zhao, X. L. 2007 The slx5-slx8 complex affects sumoylation of DNA repair proteins and negatively regulates recombination. *Molecular and Cellular Biology* 27, 6153-6162.

Burt, A. & Koufopanou, V. 2004 Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Current opinion in genetics & development* 14, 609-615.

Burt, A. & Trivers, R. L. 2006 *Genes In Conflict : The Biology Of Selfish Genetic Elements*: The Belknap Press, Harvard University.

Butler, G., Kenny, C., Fagan, A., Kurischko, C., Gaillardin, C. & Wolfe, K. H. 2004 Evolution of the MAT locus and its Ho endonuclease in yeast species. *Proceedings of the National Academy of Sciences of the United States of America* 101, 1632-1637.

Calcott, P. H. & Rose, A. H. 1982 Freeze-Thaw and Cold-Shock Resistance of *Saccharomyces cerevisiae* as Affected by Plasma Membrane Lipid Composition. *Journal of General Microbiology* 128, 549-555.

Carter, D. M. 2005 Saccharomyces genome resequencing project: http://www.sanger.ac.uk/Teams/Team118/sgrp/æ: Wellcome Trust Sanger Institute.

Charlesworth, B. 2006 The evolutionary biology of sex. *Current Biology* 16, R693-R695.

Charlesworth, D., Morgan, M. T. & Charlesworth, B. 1993 Mutations accumulation in finite outbreeding and inbreeding populations *Genetical Research* 61, 39-56.

Chen, X. J., Cong, Y. S., Wesolowskilouvel, M., Li, Y. Y. & Fukuhara, H. 1992 Characterisation of a Circular Plasmid from the Yeast *Kluyveromyces waltii*. *Journal of General Microbiology* 138, 337-345.

Chen, X. J., Saliola, M., Falcone, C., Bianchi, M. M. & Fukuhara, H. 1986 Sequence Organisation of the Circular Plasmid pKD1 from the Yeast *Kluyveromyces drosophilarum*. *Nucleic Acids Research* 14, 4471-4481.

Chen, X. L., Reindle, A. & Johnson, E. S. 2005 Misregulation of 2 mu m circle copy number in a SUMO pathway mutant. *Molecular and Cellular Biology* 25, 4311-4320.

Coluccio, A. & Neiman, A. M. 2004 Interspore bridges: a new feature of the *Saccharomyces cerevisiae* spore wall, vol. 150, pp. 3189-3196.

Crow, J. F. 1988 The ultraselfish gene. *Genetics* 118, 389-391.

Cui, H., Ghosh, S. K. & Jayaram, M. 2009 The selfish yeast plasmid uses the nuclear motor Kip1p but not Cin8p for its localization and equal segregation. *Journal of Cell Biology* 185, 251-264.

Dahlberg, C. & Chao, L. 2003 Amelioration of the cost of conjugative plasmid carriage in Eschericha coli K12. *Genetics* 165, 1641-1649.

Danhash, N., Gardner, D. C. J. & Oliver, S. G. 1991 Heritable damage to yeast caused by transformation. *Bio-Technology* 9, 179-182.

Dionisio, F., Conceicao, I. C., Marques, A. C. R., Fernandes, L. & Gordo, I. 2005 The evolution of a conjugative plasmid and its ability to increase bacterial fitness. *Biology Letters* 1, 250-252.

Dobson, M. J., Pickett, A. J., Velmurugan, S., Pinder, J. B., Barrett, L. A., Jayaram, M. & Chew, J. S. K. 2005 The 2 mu m plasmid causes cell death in Saccharomyces cerevisiae with a mutation in Ulp1 protease. *Molecular and cellular biology* 25, 4299-4310.

Dujon, B. 1981 Mitochondrial Genetics and Functions. In *The Molecular Biology of the Yeast Saccharomyces,*, pp. 505-635: Cold Spring Harbour Lab. Press.

Elena, S. F. & de Visser, J. A. G. M. 2003 Environmental stress and the effects of mutation. *J Biol* 2, 12.

Feagin, J. E., Werner, E., Gardner, M. J., Williamson, D. H. & Wilson, R. J. M. 1992 Homologies between the contiguous and fragmened ribosomal -RNAs of the 2 Plasmodium-falciparium extrachromosomal DNAs are limited to core sequences. *Nucleic Acids Research* 20, 879-887.

Frank, A. C. & Wolfe, K. H. 2009 Evolutionary Capture of Viral and Plasmid DNA by Yeast Nuclear Chromosomes. *Eukaryotic Cell* 8, 1521-1531.

Frank, S. A. 1996 Models of parasite virulence. *Quarterly Review of Biology* 71, 37-78.

Futcher, A. & Cox, B. S. 1984 Copy number and the stability of 2-mu m circle -based artifical plasmids of Saccharomyces cerevisiae. *Journal of bacteriology* 157, 283-290.

Futcher, A. B. & Cox, B. S. 1983 Maintenance of the 2-Mu-M Circle Plasmid in Populations of *Saccharomyces cerevisiae*. *Journal of Bacteriology* 154, 612-622.

Futcher, A. B. 1986 Copy number amplification of the 2-mu-m circle plasmid of Saccharomyces-cerevisiae. *Journal of Theoretical Biology* 119, 197-204.

Futcher, A. B., Reid, E. & Hickey, D. A. 1988 Maintenance of the 2pm Circle Plasmid of *Saccharomyces cerevisiae* by Sexual Transmission: An Example of a Selfish DNA. *Yeast* 4, 27-40.

Gerbaud, C. & Guerineau, M. 1980 2-Mu-M Plasmid Copy Number in Different Yeast Strains and Repartition of Endogenous and 2-Mu-M Chimeric Plasmids in Transformed Strains. *Current Genetics* 1, 219-228.

Ghosh, S. K., Huang, C. C., Hajra, S. & Jayaram, M. 2010 Yeast cohesin complex embraces 2 micron plasmid sisters in a tri-linked catenane complex. *Nucleic Acids Research* 38, 570-584.

Goddard, M. R., Charles, H., Godfray, J. & Burt, A. 2005 Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* 434, 636-640.

Goddard, M., Greig, D. & Burt, A. 2001 Outcrossed sex allows a selfish gene to invade yeast populations. *Proceedings of the Royal Society B: Biological Sciences* 268, 2537-2542.

Greig, D. 2009 Reproductive isolation in Saccharomyces. *Heredity* 102, 39-44.

Haddrill, P. R., Halligan, D. L., Tomaras, D. & Charlesworth, B. 2007 Reduced efficacy of selection in regions of the Drosophila genome that lack crossing over. *Genome Biology* 8, 9.

Hajra, S., Ghosh, S. K. & Jayaram, M. 2006 The centromere-specific histone variant Cse4p (CENP-A) is essential for functional chromatin architecture at the yeast 2-{micro}m circle partitioning locus and promotes equal plasmid segregation, vol. 174, pp. 779-790.

Hartl, D. L. & Clark, A. G. 1997 *Principles of Population Genetics*: Sinauer Associates Inc.

Hartley, J. L. & Donelson, J. E. 1980 Nucleotide-Sequence of the Yeast Plasmid. *Nature* 286, 860-865.

Hastings, I. M. 1992 Population genetic aspects of deleterious cytoplasmic genomes and their effect on the evolution of sexual reproduction. *Genetical Research* 59, 215-225.

Hickey, D. A. 1982 Selfish DNA: A sexually-transmitted nuclear parasite, vol. 101, pp. 519-531.

Hickey, D. A. 1993 Molecular Symbionts and the Evolution of Sex, vol. 84, pp. 410-414.

Holm, C. 1982 Sensitivity to the Yeast Plasmid 2-Mu DNA Is Conferred by the Nuclear Allele Nib-1. *Molecular and Cellular Biology* 2, 985-992.

Huber, D. H. & Rustchenko, E. 2001 Large circular and linear rDNA plasmids in Candida albicans. *Yeast* 18, 261-272.

Hurst, G. D. D. & Werren, J. H. 2001 The role of selfish genetic elements in eukaryotic evolution. *Nature Reviews Genetics* 2, 597-606.

Hurst, L. D. & Hamilton, W. D. 1992 Cytoplasmic fusion and the nature of sexes. *Proceedings of the Royal Society of London Series B-Biological Sciences* 247, 189-194.

Hurst, L. D. 1992 Intragenomic conflict as an evolutionary force. *Proceedings of the Royal Society of London Series B-Biological Sciences* 248, 135-140.

Hurst, L. D. 1995 Selfish genetic elements and their role in evolution - The evolution of sex and some of what that entails. *Philosophical Transactions of the Royal Society B-Biological Sciences* 349, 321-332.

Hutter, A. & Oliver, S. G. 1998 Ethanol production using nuclear petite yeast mutants. *Applied Microbiology and Biotechnology* 49, 511-516.

Jackson, C. L. & Hartwell, L. H. 1990 Courtship in Saccharomyces-cerevisiae - Both cell-types choose mating partners by responding to the strongest pheromone signal. *Cell* 63, 1039-1051.

Jasnos, L., Tomala, K., Paczesniak, D. & Korona, R. 2008 Interactions between stressful environment and gene deletions alleviate the expected average loss of fitness in yeast. *Genetics* 178, 2105-2111.

Jayaram, M., Mehta, S., Uzri, D., Voziyanov, Y. & Velmurugan, S. 2004 Site-specific recombination and partitioning systems in the stable high copy propagation of the 2-micron yeast plasmid. In *Progress in Nucleic Acid Research and Molecular Biology, Vol 77*, vol. 77, pp. 127-172. San Diego: Elsevier Academic Press Inc.

Johnson, L. J. 2007 The genome strikes back: The evolutionary importance of defence against mobile elements. *Evolutionary Biology* 34, 121-129.

Johnson, L. J., Koufopanou, V., Goddard, M. R., Hetherington, R., Schafer, S. M. & Burt, A. 2004 Population Genetics of the Wild Yeast Saccharomyces paradoxus, vol. 166, pp. 43-52.

Jones, R. N. 1991 B-chromosome drive. *American Naturalist* 137, 430-442.

Kado, C. I. 1998 Origin and evolution of plasmids. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* 73, 117-126.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. 2003 Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423, 241-254.

Knell, R. J. 1999 Sexually transmitted disease and parasite-mediated sexual selection. *Evolution* 53, 957-961.

Koufopanou, V., Hughes, J., Bell, G. & Burt, A. 2006 The spatial scale of genetic differentiation in a model organism: the wild yeast Saccharomyces paradoxus. *Philosophical Transactions of the Royal Society B-Biological Sciences* 361, 1941-1946.

Kraaijeveld, A. R., Ferrari, J. & Godfray, H. C. J. 2002 Costs of resistance in insect-parasite and insect-parasitoid interactions. *Parasitology* 125, S71-S82.

Lenski, R. E. 1988a Experimental studies of pleiotropy and epistasis in Escherichia-coli .1. Variation in competitive fitness amoung mutants resistant to virus-T4. *Evolution* 42, 425-432.

Lenski, R. E. 1988b Experimental studies of pleiotropy and epistasis in Escherichia-coli .2. Compensation for maladaptive effects associated with resistance to virus-T4. *Evolution* 42, 433-440.

Liti, G. & Louis, E. J. 2005 Yeast evolution and comparative genomics. *Annual Review of Microbiology* 59, 135-153.

Liti, G., Barton, D. B. H. & Louis, E. J. 2006 Sequence diversity, reproductive isolation and species concepts in Saccharomyces. *Genetics* 174, 839-850.

Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., Davey, R. P., Roberts, I. N., Burt, A., Koufopanou, V., Tsai, I. J., Bergman, C. M., Bensasson, D., O'Kelly, M. J. T., van Oudenaarden, A., Barton, D. B. H., Bailes, E., Ba, A. N. N., Jones, M., Quail, M. A., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R. & Louis, E. J. 2009 Population genomics of domestic and wild yeasts. *Nature* 458, 337-341.

Livingston, D. M. 1977 Inheritance of 2-Mu-M DNA Plasmid from *Saccharomyces*. *Genetics* 86, 73-84.

Lyttle, T. W. 1977 Experimental population-genetics of meiotic drive systems .1. Pseudo-Y chromosomal drive as a means of eliminating cage populations of Drosophila-melanogaster. *Genetics* 86, 413-445.

Lyttle, T. W. 1991 Segregation distorters. *Annual Review of Genetics* 25, 511-557.

Malik, H. S. & Henikoff, S. 2009 Major Evolutionary Transitions in Centromere Complexity. *Cell* 138, 1067-1082.

Marini, A., Matmati, N. & Morpurgo, G. 1999 Starvation in yeast increases non-adaptive mutation. *Current Genetics* 35, 77-81.

Martin, G., Kassen, R., Batalillon, T., MacLean, C. & Lenormand, T. 2010 The distribution of the pleiotropic fitness effects of beneficial mutations: theory and application to antibiotic resistance in Pseudomonas. *PLoS Biology* In review.

McDonald, J. H. & Kreitman, M. 1991 Adaptive protein evolution at the ADH locus in Drosophila. *Nature* 351, 652-654.

Mead, D. J., Gardner, D. C. J. & Oliver, S. G. 1986 The Yeast 2-Mu Plasmid - Strategies for the Survival of a Selfish DNA. *Molecular & General Genetics* 205, 417-421.

Mead, D. J., Gardner, D. C. J. & Oliver, S. G. 1987 Phenotypic Differences between Induced and Spontaneous 2-Mu-Plasmid-Free Segregants of *Saccharomyces cerevisiae*. *Current Genetics* 11, 415-418.

Mehta, S., Yang, X. M., Chan, C. S., Dobson, M. J., Jayaram, M. & Velmurugan, S. 2002 The 2 micron plasmid purloins the yeast cohesin complex: a mechanism for coupling plasmid partitioning and chromosome segregation? *Journal of Cell Biology* 158, 625-637.

Modi, R. I. & Adams, J. 1991 Coevolution in Bacterial-Plasmid Populations. *Evolution* 45, 656-667.

Moller, A. P. 1996 Effects of host sexual selection on the population biology, of parasites. *Oikos* 75, 340-344.

Mortimer, R. K., Romano, P., Suzzi, G. & Polsinelli, M. 1994 Genome renewal - A new phenomenon revealed from a genetic-study of 43 strains of Saccharomyces-cerevisae derived from natural fermentation of grape musts. *Yeast* 10, 1543-1552.

Murray, J. A. H., Cesareni, G. & Argos, P. 1988 Unexpected divergence and molecular coevolution in yeast plasmids. *Journal of Molecular Biology* 200, 601-607.

Murray, J. A. H., Scarpa, M., Rossi, N. & Cesareni, G. 1987 Antagonistic Controls Regulate Copy Number of the Yeast 2-Mu Plasmid. *Embo Journal* 6, 4205-4212.

Nakamura, Y., Gojobori, T. & Ikemura, T. 2000 Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research* 28, 292-292.

Nakayashiki, T., Kurtzman, C. P., Edskes, H. K. & Wickner, R. B. 2005 Yeast prions URE3 and PSI+ are diseases. *Proceedings of the National Academy of Sciences of the United States of America* 102, 10575-10580.

Nowak, M. A. & May, R. M. 1994 Superinfection and the Evolution of Parasite Virulence. *Proceedings of the Royal Society B: Biological Sciences* 255, 81-89.

Partridge, L. & Hurst, L. D. 1998 Sex and conflict. *Science* 281, 2003-2008.

Patnaik, P. R. 2000 An evaluation of models for the effect of plasmid copy number on bacterial growth rate. *Biotechnology Letters* 22, 1719-1725.

Puigbo, P., Bravo, I. G. & Garcia-Vallve, S. 2008 CAIcal: A combined set of tools to assess codon usage adaptation. *Biology Direct* 3, 8.

Rank, G. H., Xiao, W., Kolenovsky, A. & Arndt, G. 1988 FLP recombinase induction of the breakage-fusion-bridge cycle and gene conversion in Saccharomyces-cerevisiae. *Current Genetics* 13, 273-281.

Rebollo, R., Horard, B., Hubert, B. & Vieira, C. 2010 Jumping genes and epigenetics: Towards new species. *Gene* 454, 1-7.

Rogers, D. W. & Greig, D. 2009 Experimental evolution of a sexually selected display in yeast. *Proceedings of the Royal Society B-Biological Sciences* 276, 543-549.

Rose, A. B. & Broach, J. R. 1990 Propagation and expression of clones genes in yeast 2-mu-m circle based vectors. *Methods in enzymology* 185, 234-279.

Ruderfer, D. M., Pratt, S. C., Seidel, H. S. & Kruglyak, L. 2006 Population genomic analysis of outcrossing and recombination in yeast. *Nature Genetics* 38, 1077-1081.

Schaack, S., Pritham, E. J., Wolf, A. & Lynch, M. 2010 DNA transposon dynamics in populations of *Daphnia pulex* with and without sex. *Proceedings of the Royal Society B-Biological Sciences* awaiting publication.

Sharp, P. M. & Li, W. H. 1987 The rate of synomymous substitution in Enterobacterial genes is inversely related to codon usage bias. *Molecular Biology and Evolution* 4, 222-230.

Sharp, P. M., Emery, L. R. & Zeng, K. 2010 Forces that influence the evolution of codon bias. *Philosophical Transactions of the Royal Society B-Biological Sciences* 365, 1203-1212.

Sleep, D., Finnis, C., Turner, A. & Evans, L. 2001 Yeast 2 mu m plasmid copy number is elevated by a mutation in the nuclear gene UBC4. *Yeast* 18, 403-421.

Sniegowski, P. D., Dombrowski, P. G. & Fingermana, E. 2002 *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Research* 1, 299-306.

Sokal, R. R. & Rohlf, F. J. 1981 *Biometry: The Principles and Practice of Statistics in Biological Research*: W. H. Freeman and Company.

Sorensen, M. A. & Pedersen, S. 1991 Absolute invivo translation rates of individual codons in Escherichia coli- The 2 glutamic acid codons GAA and GAG are translated with a threefold difference in rate *Journal of Molecular Biology* 222, 265-280.

Springman, R., Keller, T., Molineux, I. J. & Bull, J. J. 2009 Evolution at a high imposed mutation rate: adaptation obscures the load in phage T7. *Genetics* 184, 221-32.

Steels, H., Bond, C. J., Collins, M. D., Roberts, I. N., Stratford, M. & James, S. A. 1999 Zygosaccharomyces lentus sp. nov., a new member of the yeast genus Zygosaccharomyces Barker. *International Journal of Systematic Bacteriology* 49, 319-327.

Stoebel, D. M., Deant, A. M. & Dykhuizen, D. E. 2008 The cost of expression of Escherichia coli lac operon proteins is in the process, not in the products. *Genetics* 178, 1653-1660.

Storici, F. & Bruschi, C. V. 1997 Molecular engineering with the FRT sequence of the yeast 2 mu m plasmid: [cir degrees] segregant enrichment by counterselection for 2 mu m site-specific recombination. *Gene* 195, 245-255.

Taylor, D. R., Zeyl, C. & Cooke, E. 2002 Conflicting levels of selection in the accumulation of mitochondrial defects in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* 99, 3690-3694.

Tohe, A. & Utatsu, I. 1985 Physical and functional structure of a yeadt plasmi, PSB3, isolated form Zygosaccharomyces-bisporus. *Nucleic Acids Research* 13, 4267-4283.

Tohe, A., Tada, S. & Oshima, Y. 1982 2-um DNA-Like Plasmids in the Osmophilic Haploid Yeast *Saccharomyces rouxii*. *Journal of Bacteriology* 151, 1380-1390.

Torres, E. M., Sokolsky, T., Tucker, C. M., Chan, L. Y., Boselli, M., Dunham, M. J. & Amon, A. 2007 Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* 317, 916-924.

Tsai, I. J., Bensasson, D., Burt, A. & Koufopanou, V. 2008 Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. *Proceedings of the National Academy of Sciences of the United States of America* 105, 4957-4962.

Tsalik, E. L. & Gartenberg, M. R. 1998 Curing Saccharomyces cerevisiae of the 2 micron plasmid by targeted DNA damage. *Yeast* 14, 847-852.

Tuller, T., Waldman, Y. Y., Kupiec, M. & Ruppin, E. 2010 Translation efficiency is determined by both codon bias and folding energy. *Proceedings of the National Academy of Sciences of the United States of America* 107, 3645-3650.

Turner, P. E. & Chao, L. 1998 Sex and the Evolution of Intrahost Competition in RNA Virus {phi}6, vol. 150, pp. 523-532.

Utatsu, I., Imura, T. & Tohe, A. 1988 Possible gene interchange between plasmid and chromosome in yeast. *Yeast* 4, 179-190.

Utatsu, I., Sakamoto, S., Imura, T. & Tohe, A. 1987 Yeast Plasmids Resembling 2-Mu-M DNA - Regional Similarities and Diversities at the Molecular-Level. *Journal of Bacteriology* 169, 5537-5545.

Valizadeh, P. & Crease, T. J. 2008 The association between breeding system and transposable element dynamics in Daphnia pulex. *Journal of Molecular Evolution* 66, 643-654.

Velmurugan, S., Mehta, S. & Jayaram, M. 2003 Selfishness in moderation: Evolutionary success of the yeast plasmid. In *Current Topics in Developmental Biology, Vol 56*, vol. 56, pp. 1-+.

Volkert, F. C., Wilson, D. W. & Broach, J. R. 1989 Deoxyribonucleic acid plasmids in yeasts, vol. 53, pp. 299-317.

Watve, M. M., Dahanukar, N. & Watve, M. G. 2010 Sociobiological Control of Plasmid Copy Number in Bacteria. *Plos One* 5, 8.

Xiao, W. & Rank, G. H. 1993 Polymorphism of 2-Mu-M Plasmids in Industrial Strains of *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology* 39, 80-86.

Xiao, W. & Rank, G. H. 1996 The 2um plasmid of laboratory yeast strains is a type-1/type-2 hybrid. *Yeast* 12, 809-813.

Xiao, W., Pelcher, L. E. & Rank, G. H. 1991 Evidence for cis acting and trans acting element coevolution of the 2-mu-m circle genome in *Saccharomyces cerevisiae*. *Journal of molecular evolution* 32, 145-152.

Xiao, W., Pelcher, L. E. & Rank, G. H. 1991a DNA Sequence Divergence and Functional Conservation at the STB Locus of Yeast 2μm Circle Variants. *Journal of Bacteriology* 173, 1181-1186.

Xiao, W., Pelcher, L. E. & Rank, G. H. 1991b Evidence for cis acting and trans acting element coevolution of the 2-mu-m circle genome in *Saccharomyces cerevisiae*. *Journal of molecular evolution* 32, 145-152.

Xie, Y., Pelcher, L. E. & Rank, G. H. 1994 Chimeric Evolution of the 2 mu m Genome in *Saccharomyces cerevisiae*. *Journal of molecular evolution* 38, 363-368.

Xiong, L., Chen, X. L., Silver, H. R., Ahmed, N. T. & Johnson, E. S. 2009 Deficient SUMO Attachment to Flp Recombinase Leads to Homologous Recombination-dependent Hyperamplification of the Yeast 2 mu m Circle Plasmid. *Molecular Biology of the Cell* 20, 1241-1251.

Xu, Z. J., Mitsui, K., Motizuki, M., Yaguchi, S. & Tsurugi, K. 1999 The DLP1 mutant of the yeast Saccharomyces cerevisiae with an increased copy number of the 2 mu plasmid shows a shortened lifespan. *Mechanisms of Ageing and Development* 110, 119-129.

Zakian, V. A., Brewer, B. J. & Fangman, W. L. 1979 Replication of each copy of the yeast 2-micron-DNA plasmid occurs during the S-phase. *Cell* 17, 923-934.

Zeyl, C. & DeVisser, J. 2001 Estimates of the rate and distribution of fitness effects of spontaneous mutation in Saccharomyces cerevisiae. *Genetics* 157, 53-61.

Zeyl, C., Bell, G. & Green, D. M. 1996 Sex and the spread of retrotransposon Ty3 in experimental populations of Saccharomyces cerevisiae. *Genetics* 143, 1567-1577.