

# Artificial chromosomes: ideal vectors?

William R.A. Brown, P. Joe Mee and Ming Hong Shen

Artificial chromosomes are DNA molecules of predictable structure, which are assembled *in vitro* from defined constituents that behave with the properties of natural chromosomes. Artificial chromosomes were first assembled in budding yeast and have since been useful in many aspects of yeast genetics. Several attempts have been made at building artificial chromosomes in mammals, although these have been met with limited success. Consequently, mini-chromosomes of defined structure have been developed to address questions regarding mammalian chromosome function and for biotechnological applications. Here we review progress in these areas and consider how it influences plans to build artificial chromosomes in plants and parasites.

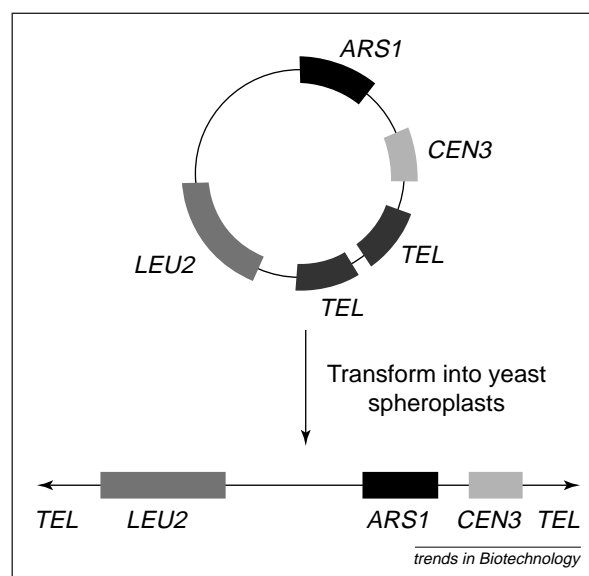
**A**rartificial chromosomes were first constructed in the budding yeast *Saccharomyces cerevisiae*<sup>1</sup>. A circular plasmid consisting of a yeast centromere, an origin of replication, a selectable marker gene and a palindromic arrangement of two stretches of telomeric DNA was assembled by conventional recombinant DNA techniques and introduced into yeast by spheroplast transformation where it resolved into a simple linear molecule (Fig. 1). Linear constructs, 50 kb in length, containing a centromere, an origin of replication and two telomeres were replicated and segregated at mitosis with ~99% accuracy, and were retained in dividing cultures for at least 20 generations<sup>1</sup>.

Three factors contributed to the success of these experiments: (1) each of the three *as*-acting sequences necessary for yeast chromosome function<sup>2</sup> had been characterized before these experiments and could be isolated on small fragments of DNA that could be easily manipulated; (2) the yeast could be transformed efficiently and grew sufficiently rapidly for controlled experiments to be performed quickly and easily; and (3) perhaps most importantly, natural yeast chromosomes are sufficiently small that it was possible to assemble a molecule that was large enough to function as a yeast chromosome with only minor modifications of conventional recombinant DNA techniques. The ability to assemble YACs suggested the possibility of assembling artificial chromosomes in other species, particularly plants and mammals. However, this goal has not been reached and the difficulties reflect the fact that, on the whole, plants and animals have much longer and more complex chromosomes than yeast. Despite limited practical achievements of the work to date, however, it is now known with reasonable certainty how large a chromosome needs to be if it is to segregate accurately at cell division in human cells<sup>3,4</sup> and what sequences normally function as centromeres on human chromosomes. Structurally defined mini-chromosome vectors have also been developed for the mammalian

germ line<sup>5,6</sup> and although these are not assembled *ab initio*, they might be useful for introducing long tracts of DNA into the germ line and for studying the genetics of mammalian chromosome transmission.

## Sequences necessary for building mammalian artificial chromosomes

The experiments carried out on yeast artificial chromosomes suggested that three different types of *as*-acting DNA sequence would be required to build a mammalian artificial chromosome: (1) telomeres, (2) specific origins of replication and (3) a centromere. In yeast, molecules that were one fifth the size of the smallest natural chromosomes segregated with greater than 99% accuracy<sup>1</sup> and it therefore seemed reasonable to anticipate that an artificial mammalian chromosome would need to be ~10 Mb in length if it were to be mitotically stable.



**Figure 1**

The construction of artificial chromosomes in the yeast *Saccharomyces cerevisiae*. A plasmid containing a selectable marker gene (LEU2), a centromere (CEN3), a yeast origin of replication (ARS1) and two telomeric sequences was assembled in *Escherichia coli* and introduced into *S. cerevisiae* by spheroplast transformation. After transformation, the telomeric sequences resolved to yield a linear derivative of the starting plasmid.

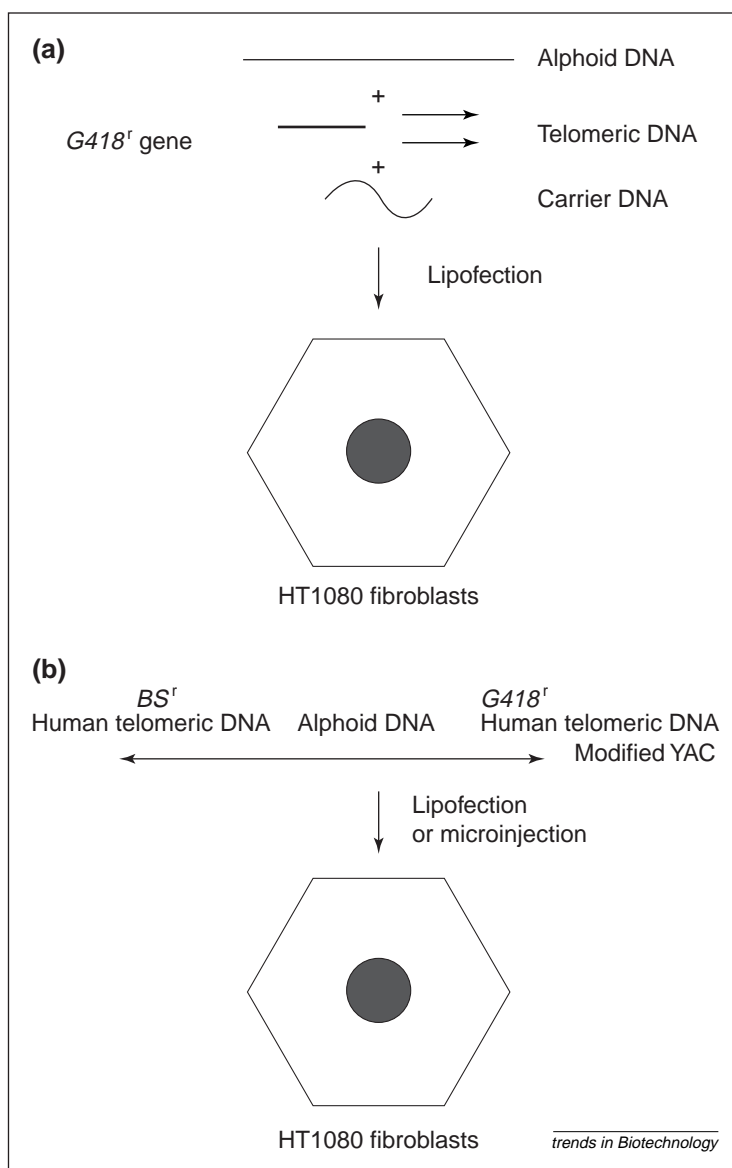
W.R.A. Brown ([wrab@bioch.ox.ac.uk](mailto:wrab@bioch.ox.ac.uk)) is at the Institute of Genetics, University of Nottingham, Queen's Medical Centre, Nottingham, UK NG7 2UH. P.J. Mee is at the University of Edinburgh, Centre for Genome Research, King's Buildings, West Mains Road, Edinburgh, UK EH9 3JQ. M.H. Shen is at the Cancer Research Campaign Chromosome Molecular Biology Group, Biochemistry Department, South Parks Road, Oxford, UK OX1 3QU.

In mammals and birds, the telomeric DNA has been isolated as  $(TTAGGG)_n$ , and arrays of  $(TTAGGG)_n$  that are ~1 kb in length have been shown to function with ~70% efficiency when introduced into human cells grown in the laboratory<sup>7-9</sup>. Thus, telomeric DNA is the most well understood and easily manipulatable of the three sequences necessary for mammalian chromosome function. A wide range of different sequences can function as origins of replication in human fibroblast cells, and it is unclear whether the construction of a mammalian artificial chromosome requires the provision of a specific origin sequence or if this role could be assumed by one of the other *cis*-acting sequences or by a selectable marker gene. Centromeres pose significant problems to the construction of mammalian artificial chromosomes for two reasons: (1) the DNA sequences required for centromere function are poorly defined and (2) the sequences that have been shown to confer centromere function are difficult to manipulate.

Studies of natural and engineered rearranged chromosomes indicate that alphoid DNA is the functional centromeric DNA on most human chromosomes. The existence of stable marker chromosomes that lack alphoid DNA demonstrates that non-alphoid DNA sequences will function as centromeres<sup>10</sup> but the fact that, despite considerable variation in amount, all human centromeres contain some alphoid DNA suggests strongly that centromeres form most-readily over alphoid DNA. The amount of alphoid DNA required for accurate segregation is not known but mini-chromosomes have been identified with only 90 kb of alphoid DNA, which segregates accurately at mitosis (J.W. Yang, pers. commun.). Although 90 kb tracts of alphoid DNA are stable and can be manipulated in yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs), it is not clear whether or not this is enough alphoid DNA to initiate formation of a new centromere. Another problem is that mammalian chromosomes, even small ones, are much larger than those that can be conveniently manipulated by cloning systems. The smallest human chromosome, chromosome 22, is approximately 50 Mb in length; if it is assumed that the relationship between size and stability scales between yeast and humans then, in humans, chromosomes smaller than 10 Mb in length would be anticipated to be unstable. In fact, mammalian mini-chromosomes ~5 Mb in length are stably maintained in culture in somatic cells<sup>3</sup>, and it is only at ~2.5 Mb that evidence of instability is seen<sup>4</sup>. Although this is larger than can be conveniently isolated in most cloning systems, it is possible to use homologous recombination in *S. cerevisiae* to link YACs<sup>11</sup> and to assemble molecules as large as 2.3 Mb, and then introduce these into human fibroblasts by transfection<sup>12</sup>. A major limitation of this approach, however, might be that these large YACs are unstable in yeast and have a limited capacity for cloning the long tracts of repeated sequences that are necessary for centromere function.

### Building mammalian artificial chromosomes

Two protocols have demonstrated that it is possible to assemble fully functioning artificial chromosomes from cloned DNA in mammalian cells. In the first protocol, alphoid DNA was cloned in bacterial artificial chromosomes and then co-transfected with telomeric



**Figure 2**

Construction of artificial chromosomes in human fibroblasts. **(a)** Human alphoid DNA cloned in bacterial artificial chromosomes was mixed *in vitro* with a selectable marker gene conferring resistance to the antibiotic G418, telomeric DNA and, in some cases, human carrier DNA and was then introduced into human HT1080 fibroblasts by lipofection. In a small proportion of the stably transfected clones, accurately segregating episomal structures were formed exclusively from these constituents. **(b)** A YAC containing 90 kb of human alphoid DNA from the centromere of chromosome 21 was modified by homologous recombination in yeast to contain human telomeric sequences and selectable marker genes conferring resistance to the antibiotics G418 and blasticidin, and then introduced into HT1080 fibroblasts by either lipofection or microinjection. In ~50% of the stably transfected clones stably replicating, accurately segregating episomes were formed exclusively from the YAC DNA.

DNA into human HT1080 fibrosarcoma cells<sup>13</sup> (Fig. 2a). In some transfections, human or mouse genomic DNA was included in the mixture as a carrier. When the mixture included alphoid DNA, telomeric DNA and human genomic DNA, mini-chromosomes containing the transfected alphoid sequences were seen in nine out of 27 transfected clones. These mini-chromosomes were characterized cytogenetically, which showed that they all had complex structures. Many included sub-telomeric sequences from natural chromosomes,

including alphoid DNA from the acrocentric chromosomes (13,14,15,21,22), but one consisted exclusively of input DNA and was linear. This mini-chromosome was stably maintained in the absence of any applied selection. In the absence of any evidence for the incorporation of sequences from the host chromosome, it was concluded that this mini-chromosome and its centromere were formed *de novo* from input DNA. The fact that a mixture of DNA sequences was used to transform the cells meant that a rigorous comparison between the structures of the input DNA and the mini-chromosome was impractical, but the observation that the mini-chromosome was several mega bp in size suggested that complex rearrangements had occurred during the process of mini-chromosome formation.

A better-defined system that has provided important evidence for the role of alphoid DNA in centromere function was described by Ikeno *et al.*<sup>14</sup> who modified YACs containing ~100 kb of alphoid DNA from the centromere of chromosome 21 by retrofitting them with human telomeric DNA and marker genes. The YACs were then purified and introduced into human HT1080 cells by lipofection or microinjection (Fig. 2b). In ~30% of the stably transfected clones, the DNA had assembled into accurately segregating mini-chromosomes that were shown cytogenetically to be free of contaminating host sequences. Restriction analysis suggested that the mini-chromosomes consisted exclusively of concatamers of the input alphoid DNA, and fluorescence *in situ* hybridization analysis with telomeric DNA indicated that they were linear. Together, these results implied that the alphoid DNA was functioning as the centromere. In order to address this question, a YAC that contained a second alphoid sequence with less homogeneous tandem repeats and no binding sites for centromere protein B was assayed for artificial chromosome formation. This alphoid DNA, which is close to (but not at) the functional centromere, failed to seed mini-chromosome formation, demonstrating sequence specificity in the ability of particular sequences to initiate centromere formation. Although the cytogenetic studies suggested that the mini-chromosomes were just a few mega bp in size, they could not be size fractionated by pulsed field gel electrophoresis. It might be that the mini-chromosomes are circular and the telomeric DNA was not functional, or the mini-chromosomes might be structurally unstable as a result of unequal sister-chromatid exchange and thus will be heterogeneous in size.

#### Artificial chromosomes and gene therapy

The systems described by Harrington *et al.*<sup>13</sup> and Ikeno *et al.*<sup>14</sup> have several limitations that make it unlikely that they will ever be used for therapeutic gene delivery. What are these limitations and can they be addressed experimentally?

The first, and most important, problem is that the strategies of Harrington *et al.* and Ikeno *et al.* generated mini-chromosomes with no predictable relationship to the input DNA. Therefore, any gene that might be incorporated into this system would most likely be in an unpredictable sequence environment. But why does the DNA rearrange? There are two possible explanations: (1) it might need to assemble a structure that is large enough to be stable as a chromosome; or (2)

concatamerization might be required for efficient centromere assembly. Recent experiments suggest that it is possible to build a YAC as large as 2.3 Mb (Ref. 11) and introduce it intact into human fibroblasts<sup>12</sup>; consequently, it might be possible to answer this question.

The second problem is that although artificial chromosomes form following transfection, rearrangements frequently occur when the transfected DNA fails to form a new chromosome. Telomeric sequences often seed the formation of new telomeres when they are integrated into host chromosomes and this, in turn, leads to the loss of large segments of chromosomal DNA distal to the integration site<sup>8</sup>. The artificial chromosome construction systems are therefore powerful mutagens.

The third problem is that the absolute frequency of artificial chromosome formation is very low; mini-chromosomes formed at a frequency of  $\sim 5 \times 10^{-5}$  per transfected cell when cationic lipids were used as a transfection reagent.

The fourth problem is that artificial chromosomes have been shown to form in only one human cell type: HT1080 fibrosarcoma cells. There have been no reports of artificial chromosome formation in any other human cell lines or in cell lines from other mammalian species. In experiments in which YACs containing human alphoid DNA were introduced into mouse LA-9 cells, extra-chromosomal structures were formed with the characteristics of double minute chromosomes that segregated randomly<sup>15</sup>. The observation that human telomeric DNA fails to function when it is introduced into primary fibroblast cells<sup>8</sup> might indicate that MAC formation is limited to cells that contain telomerase. If this is so, it will seriously constrain the application of this technology.

Despite these difficulties, both systems successfully showed that it is possible to assemble a fully functioning mammalian chromosome starting with cloned DNA. Thus, regardless of the outcome of the goal of building a useful artificial chromosome, these pioneering and difficult experiments have established that the mechanism of centromere assembly in vertebrate cells is a problem that can be addressed experimentally.

#### Artificial chromosomes in plants and human pathogens

Although most of the effort to date has focused on building mammalian artificial chromosomes, the idea that artificial chromosomes might be useful in other species has existed since the original yeast experiments. The near-completion of the *Arabidopsis* and rice genome-sequencing projects, and the success of the relatively limited genetic modification of crop plants has prompted the formulation of specific plans for plant artificial chromosomes<sup>16</sup>. The proposed advantages of plant artificial chromosomes are:

- (1) they would enable the rapid identification of genes in crop plants that complement genetically well-characterized *Arabidopsis* mutants;
- (2) they would permit many genes to be introduced simultaneously into plants in order to engineer a complex feature of a plant, such as the amino acid content of its seeds; and
- (3) they would help identify genes conferring semi-dominant traits, such as disease resistance.

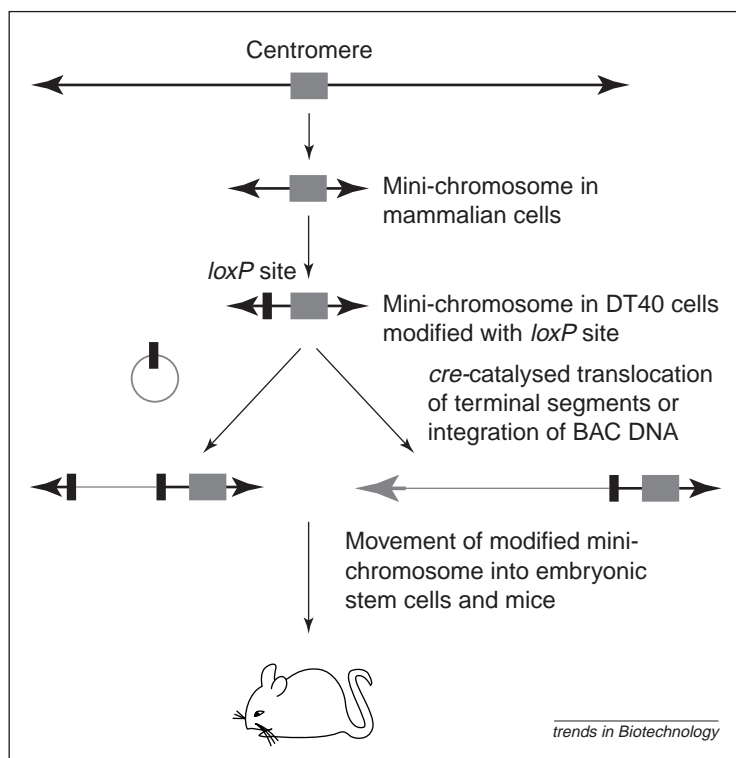
The approach that has been proposed is one that is similar in principle to that used by Ikeno *et al.*<sup>14</sup> for developing human artificial chromosomes. Clearly, the question that this experimental approach raises is: will these experiments experience the many problems encountered during the experiments aimed at building mammalian artificial chromosomes? It is impossible to know until the experiments have been done, but because both plant and human chromosomes are large and contain tandemly repeated centromeric DNA, it would seem reasonable to suspect that similar problems will arise, and it would therefore seem worthwhile to explore other approaches.

A second organism in which it would be useful to develop artificial chromosomes is the malarial parasite *Plasmodium falciparum*. Candidate centromeric sequences have been identified in this organism and shown to be only a few kb in length<sup>17</sup>. However, the DNA is >95% A+T and can only be cloned in short (~200 bp) stretches in *Escherichia coli*. *Plasmodium* chromosomes are from 400 kb to 3 Mb in size, and thus if a cloning host that is tolerant of A+T rich DNA could be found, it should be relatively easy to assemble molecules large enough to function as artificial chromosomes in this parasite using approaches similar to those used to build artificial chromosomes in yeast. Such chromosomes might have considerable application in cloning genes in *Plasmodium* by complementation, and for studying gene expression in *Plasmodium*.

### Mini-chromosome vector systems

A second approach to developing vectors with the properties of chromosomes has been to avoid the difficulties of assembling fully functional chromosomes of a specific structure from defined constituents and instead to focus on engineering mini-chromosomes of defined size and sequence composition, and to use these as vectors (Fig. 3). This approach relies on the use of cloned telomeric DNA as a reagent to systematically fragment mammalian chromosomes<sup>7,8</sup>, and site-specific recombinases<sup>18</sup> as reagents to target specific cloned DNA sequences to the engineered mini-chromosomes. This approach has become practical, with the ability to move mammalian mini-chromosomes into the hyper-recombinogenic chicken B lymphoid cell line DT40, wherein they can be readily modified by homologous recombination. It is known that unpaired fragments of human chromosomes can be passed through the mouse germ line<sup>20</sup> and it has recently been shown that it is possible to move a 4.5 Mb (Ref. 6) mini-chromosome of defined structure from somatic cells in culture into mouse embryonic stem cells and from there through the mouse germ line. This approach therefore permits large-scale transgenesis of the mammalian genome<sup>21</sup>.

It would be useful to have an artificial chromosome vector system for plants. It is possible to fuse mammalian cell and plant spheroplasts, and it should therefore be practical to investigate: (1) whether mammalian mini-chromosomes are stably maintained in plant cells and (2) whether the addition of plant centromeric sequences to the mammalian mini-chromosomes might improve stability. If this can be demonstrated, it might be possible to engineer a mini-chromosome in plant cells, which could function either as a target for experimentally interesting sequences cloned in BACs



**Figure 3**

Mini-chromosome vectors for the mouse germ line. A human chromosome isolated in a somatic cell hybrid is systematically fragmented by the integration of cloned telomeric DNA to yield an accurately segregating mini-chromosome of <10 Mb in size. The mini-chromosome is then moved into the hyper-recombinogenic chicken B lymphoid cell line DT40 where it is modified by the introduction of a *loxP* target site for the bacteriophage P1 recombinase *cre*. Subsequently, it can be modified either by the introduction of a BAC modified with a *loxP* site or by the translocation of a sub-terminal segment of a second human chromosome using *cre*-mediated site-specific recombination. The modified mini-chromosome can then be moved into the mouse germ line by modification of embryonic stem cells using somatic cell genetic techniques. The technology for each of these individual steps is established but they have not been combined to generate a functional system.

or as the foundation of a chromosome that would ultimately contain a whole set of genes of agricultural importance.

### Mini-chromosomes, gene expression and gene therapy

Practical considerations have been as important a motivation for building mammalian mini-chromosomes as for building artificial chromosomes. However, what exactly can engineered mini-chromosomes be used for? Engineered mini-chromosomes have three important properties: (1) they have defined structures; (2) they can, at least in principle, be engineered to contain large tracts of DNA, although this has not yet been shown experimentally; and (3) they can be passed through the vertebrate germ line. Engineered mini-chromosomes thus constitute a unique vector system for introducing large tracts of DNA into somatic cells and the germ line of experimental animals and (in principle) agricultural livestock. A major problem with their use is that the ability to move mini-chromosomes between different cell types relies on the techniques of somatic cell genetics and, in particular, on microcell transfer. These techniques are very slow and labour intensive. A second,

more fundamental, limitation of the technology is that most mammalian genes are contained within a few kb or tens of kb, and the range of interesting questions that one might wish to address using engineered mini-chromosomes is limited.

Considering the basic scientific questions that one might wish to address using engineered mini-chromosomes, one first has to address the question: what unsolved problems are there in mammalian genetics that are likely to reflect the integrated function of a large tract of DNA? The fact that the major mechanisms for mutagenesis occur within, or in the immediate vicinity of, the affected gene indicate that there are few. There are, however, several interesting human genetic disorders that might reflect long-range genome organization and mutations acting through poorly understood position effects; it would be valuable to have experimental tools to analyse such disorders. One particular example of such a disorder is facioscapulo-humeral muscular dystrophy (FSHD), which segregates as an autosomal dominant mapping to the long arm of human chromosome 4 (Ref. 19). The gene or genes responsible for FSHD have not been identified but one consistent feature is that chromosomes associated with the disease are characterized by the presence of less than ten copies of a tandemly repeated sub-telomeric DNA sequence, and the wild-type chromosomes have a variable number of copies in excess of ten (Ref. 19). Because no mutation has been associated with the disease, it has been suggested that it arises from the deregulated expression of genes in the sub-telomeric domain of 4q as a consequence of the deletion of sub-telomeric repeats. It might be possible to use a mini-chromosome vector to test this hypothesis. Thus, by mobilizing the terminal region of the long arm of human chromosome 4 onto a germ-line-competent mini-chromosome and then introducing derivative mini-chromosomes with either the wild type or the FSHD pattern of sub-telomeric repeats into the mouse germ line, one might attempt to model the disease.

A second type of problem where mini-chromosomes might be useful is in the identification of mouse genes that affect chromosome stability. Several tumour suppressor genes have been suggested to affect the accuracy of chromosome segregation at mitosis and it would be valuable to have assays for such genes. Preliminary work suggests that a 4.5 Mb mini-chromosome is slightly mitotically unstable after transfer through the mouse germ line<sup>6</sup>. It might be possible to assay this instability by introducing a coat-colour marker gene into the mini-chromosome and then using variegation of the coat colour to identify genes that affect mini-chromosome stability. A third area of applicability has been pursued by Fisher *et al.* who have developed the use of trans-chromosomal animals as models of human syndromes arising from chromosomal aneuploidy, for example, trisomy 21 or Down's syndrome<sup>22</sup>. Given the developmental differences between mouse and human, however, it remains to be seen how useful these interesting experiments will be.

It should also be practical to use mini-chromosome vectors to engineer the genomes of animals for biotechnological applications. One important example of this type of approach is the recent construction of mice containing fragments of human chromosomes with

each of the immunoglobulin heavy chain and  $\kappa$  light chain genes on a genetic background in which the mouse genes encoding each of these proteins was deleted<sup>21</sup>. This allowed the production of human antibodies in mice and the production of mouse hybridomas secreting human antibodies. The approach used in these procedures, however, relied on random fragments of human chromosomes, and thus has two important limitations: (1) the transgenic DNA is undefined and this might create problems when it comes to control when more-subtle phenotypes are sought; and (2) some of the chromosome fragments used in these experiments were of limited stability and the human centromeres on the fragments might be unable to function efficiently in the mouse cells. More experiments will be required to establish how well human centromeres function in mouse cells, and whether or not these types of experiments require a vector with a mouse centromere if they are to be generally applicable. The production of human antibodies in mice is a valuable use of large-scale transgenesis but, again, how generally applicable is this sort of technology? Ishida *et al.*<sup>21</sup> have suggested that it would be useful to humanize mice with the cluster of genes from the major histocompatibility complex or genes encoding drug-metabolizing enzymes. Large-scale transgenesis of mice might also be useful for understanding the genetic basis of interspecific differences. It would be interesting to know, for example, how extensively one could substitute the mouse genome with rat DNA and whether or not one could use this to identify the genetic basis of the more-sophisticated behaviour that is shown by rats.

An important reason given for building artificial chromosomes has been to develop vectors for gene therapy. However, is there any chance that engineered mini-chromosomes could be useful in gene therapy? They have a critical advantage over artificial chromosomes constructed from cloned constituents in that they have defined structures. Thus, genes present on these mini-chromosomes can be assured of being present in a defined and controllable sequence environment. They are also autonomous and thus avoid problems of insertional mutagenesis that are seen with many viral vectors. Mini-chromosome vectors have a potentially large sequence capacity and thus, in principle, could be useful in gene therapy. The major factor limiting this potential is the low efficiency with which mini-chromosomes can be transferred between different cell types. It is difficult at present to see how this could be improved. However, if mini-chromosomes could be introduced into human embryonic stem cells and the resulting transfectants cloned and propagated then the low transfer efficiency might not be a limitation to their applicability.

### Summary

It is turning out to be more difficult than anticipated to build mammalian artificial chromosomes. However, there has been a slow but systematic progress in understanding, at least empirically, many important aspects of mammalian chromosome function, both as a result of attempts to build artificial chromosomes and of studies of mammalian mini-chromosomes. It seems likely that progress will continue and ultimately different types of vector systems with a variety of advantages and

disadvantages will emerge for plants as well as mammals. These vectors are likely to be primarily useful for the information they provide regarding how the chromosomes of mammals and plants actually work, but they will also be important as reagents for the new area of large-scale transgenesis of experimental animals, agricultural livestock and crop plants.

References

- 1 Murray, A.W. *et al.* (1986) Chromosome length controls mitotic chromosome segregation in yeast. *Cell* 45, 529–536
- 2 Blackburn, E.H. and Szostak, J.W. (1984) The molecular structure of centromeres and telomeres. *Annu. Rev. Biochem.* 53, 163–194
- 3 Heller, R. *et al.* (1996) Mini-chromosomes derived from the human Y chromosome by telomere directed chromosome breakage. *Proc. Natl. Acad. Sci. U. S. A.* 93, 7125–7130
- 4 Mills, W. *et al.* (1999) Generation of an approximately 2.4 Mb human X centromere-based minichromosome by targeted telomere-associated chromosome fragmentation in DT40. *Hum. Mol. Genet.* 8, 751–761
- 5 Shen, M.H. *et al.* (1997) Human mini-chromosomes in mouse embryonal stem cells. *Hum. Mol. Genet.* 6, 1375–1382
- 6 Shen, M.H. *et al.* (1999) A structurally defined mini-chromosome vector for the mouse germ line. *Curr. Biol.* 10, 31–34
- 7 Farr, C. *et al.* (1991) Functional re-introduction of human telomeres into mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 88, 7006–7010
- 8 Barnett, M.A. *et al.* (1993) Telomere directed fragmentation of mammalian chromosomes. *Nucleic Acids Res.* 21, 27–36
- 9 Hanish, J.P. *et al.* (1994) Stringent sequence requirements for the formation of human telomeres. *Proc. Natl. Acad. Sci. U. S. A.* 91, 8861–8865
- 10 duSart, D. *et al.* (1997) A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. *Nat. Genet.* 16, 144–153
- 11 Larin, Z. *et al.* (1996) A method for linking yeast artificial chromosomes. *Nucleic Acids Res.* 24, 4192–4196
- 12 Marschall, P. *et al.* (1999) Transfer of YACs up to 2.3 Mb intact into human cells with polyethylenimine. *Gene Ther.* 6, 1634–1637
- 13 Harrington, J.J. *et al.* (1997) Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. *Nat. Genet.* 15, 345–355
- 14 Ikeno, M. *et al.* (1998) Construction of YAC-based mammalian artificial chromosomes. *Nat. Biotechnol.* 16, 431–439
- 15 Taylor, S.S. *et al.* (1996) Analysis of extrachromosomal structures containing human centromeric alphoid satellite DNA sequences in mouse cells. *Chromosoma* 105, 70–81
- 16 Somerville, C. and Somerville, S. (1999) Plant functional genomics. *Science* 285, 380–383
- 17 Bowman, S. *et al.* (1999) The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* 400, 532–538
- 18 Smith, A.J.H. *et al.* (1995) A site directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination. *Nat. Genet.* 9, 376–385
- 19 Wijmenga, C. *et al.* (1992) Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.* 2, 26–30
- 20 Tomizuka, K. *et al.* (1997) Functional expression and germ line transmission of a human chromosome fragment in chimeric mice. *Nat. Genet.* 16, 133–143
- 21 Tomizuka, K. *et al.* (2000) Double trans-chromosomal mice: maintenance of two individual human chromosome fragments containing Ig heavy and k loci and expression of fully human antibodies. *Proc. Natl. Acad. Sci. U. S. A.* 97, 722–727
- 22 Hernandez, D. *et al.* (1999) Transchromosomal mouse embryonal stem cell lines and chimeric mice that contain freely segregating fragments of human chromosome 21. *Hum. Mol. Genet.* 8, 923–933

**Do multiple databases doom the paper book?**

Genetics Databases

edited by M.J. Bishop, Academic Press, 1999.

US\$49.95 (xiv + 295 pages) ISBN 0 12 1016250

With the dawn of the genome era, there are major concerns about the storage, annotation and integration of the new and ever-expanding wealth of biological data. There are hundreds of databases and associated World Wide Web resources, which are in different formats using different notations and nomenclature, and with a wide range of redundancies, errors and outdated entries. They range from being purely archival to being theoretical or model-derived. Some are specific to a particular organism, others cover many organisms, but only for a small set of functions or pathways.

In recent years, a sense of comfort (and perhaps a false sense of security) has emerged with

many of these divergent data resources, and there are at least two reasons for this. First, many of the data sources have proved to be extremely useful. One only has to look at the successful use of the genetic sequence and structural databases in identifying the probable function of newly sequenced proteins for which no experimental measurement exists. Second, modern computers and the Internet have allowed easy search access. However, if we disregard the old computer saying, ‘garbage in, garbage out’, even the slickest of web interfaces will not save an unwary user from problems of quality or completeness in the underlying databases. Perhaps more

importantly, the current ease of use does not provide evidence for the longer-term scalability of either the data structures or their integration.

It is relatively easy for an individual researcher working on a single pathway or a family of proteins to follow a set of Internet hot links to recover and, to a large extent, integrate a diverse array of scattered information. All that is required is some patience and expert knowledge in a narrow domain of biology; such expertise allows considerable error detection and subjective selectivity. However, asking broader questions about development, evolution, physiology and cellular signaling across any wide range of species, protein families and cellular environments cannot be done in a similar manner! Thus, database and access scalability concerns result from the complexity of anticipated queries, as well as the amount of data.

