

# Chapter 15

## Genomics of Subtelomeres: Technical Problems, Solutions and the Future

Marion M. Becker and Edward J. Louis

**Abstract** Genome projects invariably are missing the subtelomeric regions due to cloning, sequencing and informatic problems. The repetitive nature of these sequences and the shared homology between different subtelomeres in most organisms preclude assembly of the regions leaving them incomplete. This is clearly an issue when a great deal of interesting biology, as seen in the previous chapters, involves the subtelomeres. In some cases, organism-specific or individual strain-specific approaches have been used to obtain material for sequencing and analysis as well as assembly. However, a more general approach applicable to most organisms has been elusive. Various cloning techniques have been tried and developed with mixed success. The most promising general approach involves individual telomere regions cloned as yeast artificial chromosomes (YACs), isolating them from all the other subtelomeres that have overlapping homology. This is time-consuming, and there is still a bottleneck in sequencing these. New sequencing technologies may solve many of the technical problems, yet there are still informatic problems with assembly because of the repetitive nature of the regions. Although progress has been made, the solution to efficient completion of subtelomeric regions will likely take combined approaches, such as deep sequencing in pedigrees to look for genetic linkage (associations) with known segregating sites near the ends of chromosomes. An efficient and cost-effective approach is needed before individual genome projects can move into population genomics involving the subtelomeres which is crucial for studies of diversity in parasites as well as in quantitative genetic studies in many organisms.

---

M. M. Becker (✉) · E. J. Louis  
Centre for Genetic Architecture of Complex Traits, Department of Genetics, University of  
Leicester, Leicester LE1 7RH, UK  
e-mail: MarionMBecker@gmail.com

E. J. Louis  
e-mail: ejl21@le.ac.uk

## 15.1 Introduction

Most genome projects, including the human genome, are incomplete as they typically are missing the subtelomeric regions. In whole genome shotgun libraries, subtelomeric sequence is frequently missing, rearranged or underrepresented, and despite enormous effort, gaps remain in the subtelomeres. This is true for the *Caenorhabditis elegans* project (Consortium 1998), the *Drosophila genome* project (Adams et al. 2000; Celniker et al. 2002), the human genome project (Riethman et al. 2001), the *Schizosaccharomyces pombe* project (Wood et al. 2002), the *Plasmodium falciparum* project (Gardner et al. 2002), the *Trypanosoma brucei* project (Berriman et al. 2005) and many others. This is not a coincidence, and the following chapter will highlight the problems and some of the solutions that have been used to close the gaps on some of these projects.

The problems historically and currently are as follows:

1. Lack of telomeric and subtelomeric clones
2. Difficulty in cloning large enough fragments to connect with genome contigs
3. Difficulty in sequencing clones
4. Difficulty in assembling sequences

Some of these difficulties have been solved in some cases, but there has been no general approach that solves all of these for a generic genome project, though for some fungi an efficient approach has been developed (Farman 2011; Farman and Leong 1995; Li et al. 2005).

### 15.1.1 Underrepresentation of Subtelomeres in Standard Libraries

In the early days of genome projects, using first-generation Sanger sequencing and shotgun, cosmid and bacterial artificial chromosome (BAC) libraries, it was clear that telomeres and subtelomeres were underrepresented in these libraries by 10- to 100-fold (Becker et al. 2004). In the case of human subtelomeres, this is a consequence of the proximity to telomeres and the lack of restriction sites used in cloning procedures (Mefford and Trask 2002), the high GC content and the presence of telomeric repeats (Costa et al. 2009). *T. brucei* subtelomeres were more than 10-fold underrepresented in BAC libraries, and all clones isolated and sequenced were incomplete and in some cases rearranged (Berriman et al. 2002, 2005). This is due to several problems, one simply being the structure of the end of the chromosome, which needs to be enzymatically processed before being ligated into a vector. In addition, the telomere repeats are unstable in *Escherichia coli*. Even in BAC libraries, the presence of inverted repeats, AT-rich sequences and Z-DNA sequence structures are extremely unstable in *E. coli* (Kouprina et al. 2003) and cannot be cloned efficiently or at all in bacteria. Interestingly, in fungal genome

projects, the telomere sequences were overrepresented in cosmid libraries yet not incorporated into assemblies due to the region being recalcitrant to shearing during shotgun cloning (Schwartz and Farman 2010).

### ***15.1.2 Problems with Mapping Subtelomere Clones Onto the Genome***

Mapping clones of subtelomeres, if they are obtained, back to the genome is difficult due to the shared homologies between subtelomeres. Inserts in the clones from various libraries are generally smaller than the large regions of shared homology between subtelomeres, precluding a direct mapping onto the core genome. This was true for many genome projects including *C. elegans* (Consortium 1998), *S. pombe* (Wood et al. 2002) and *T. brucei* (Berriman et al. 2005). Gap filling has helped some of these projects with a great deal of effort. One method for direct isolation of chromosomal fragments is PCR, which could help with some of the problems. However, the major limitation is that DNA fragments much larger than ~20 kb cannot be easily amplified due to the shearing of template sequence and the low processivity of thermostable DNA polymerases (Kouprina and Larionov 2006). If shared homology regions are large and there are gaps within several subtelomeres, then it would not be possible to uniquely amplify a given subtelomere region. For most genomes, the only cloning vectors with large enough inserts to cover the long regions of homology are BACs and yeast artificial chromosomes (YACs). BACs are problematic as described above, but YACs have other problems as described next.

### ***15.1.3 Problems with Sequencing Large Subtelomere Clones***

As will be described below, there are now good methods for obtaining large telomere containing subtelomeric clones that map back to the core of the genome, involving linear YACs. These clones are still difficult to deal with at the sequencing level for different reasons. As already mentioned, even in telomere-enriched libraries of large insert clones, the shotgun approach can result in lack of assembled sequence due to problems with shearing (Schwartz and Farman 2010). Second-generation sequencing that does not have a cloning step in a host organism overcomes much of this problem. However, there is another problem encountered with large-subtelomere-containing clones: subtelomeres cloned as YACs must be isolated away from the yeast host genome before sequencing. This usually involves separation of the YAC from the yeast chromosomes using pulsed-field gels, purification and then sequencing, either through a shotgun library or by second-generation methods without cloning in a vector (Hertz-Fowler et al. 2008). These sequence projects, both first generation and second generation, have large

amounts of host yeast genome contamination as seen in a collection of *T. brucei* subtelomere clones (Hertz-Fowler et al. 2008). Better purification methods are required to make the whole process more efficient.

#### ***15.1.4 Assembly Problems in Repetitive Regions***

The final difficulty is not a subtelomere-specific problem but one involving repeats. Most sequence assemblers have difficulties when reads map to more than one contig location precluding completion of repetitive regions and in particular the subtelomeric regions of most organisms. For fungi where the subtelomeric repetitive regions are relatively small, informatic approaches have solved this problem (Li et al. 2005), but for many projects, including yeast, with smaller genomes than most fungi, the informatic approach has not been sufficient (Liti et al. 2009, 2013). Part of the solution is through isolated individual clones away from the rest of the genome as for example in *T. brucei* (Becker et al. 2004; Hertz-Fowler et al. 2008) and the human subtelomeres (Riethman et al. 2004), but this is time-consuming and still does not help with repeats within a given subtelomere.

### **15.2 Yeast to the Rescue and Other Solutions**

Being the first eukaryotic genome project (Goffeau et al. 1996, 1997) and one of the first eukaryotic population genomics projects (Liti et al. 2009), yeast has exposed many of these problems and has led to the development of various solutions to these problems.

#### ***15.2.1 Cloning Subtelomeres to Finish Genome Projects***

Approaches have been developed for specific projects, but these are not generally applicable to all genome projects. The yeast genome project is a case in point where these technical difficulties were solved with yeast-specific techniques. Once it was recognized that the standard library approach was not going to complete the ends, as the first eukaryotic chromosome sequenced (Oliver et al. 1992) actually was not complete (Louis 1994), each telomere was marked uniquely by inserting a vector into the telomere repeats (Louis and Borts 1995). Thirty-two strains, one for each telomere, were then used to either clone the sequences adjacent to the inserted vector as a plasmid or generate long-range PCR products using the vector as a unique anchor (Louis 1995). This approach successfully tagged and allowed the sequence and assembly of every telomere

despite the shared homologies as the marked telomeres were at specific chromosome ends and the length of the clones or PCR products spanned the large regions of homology. The standard approach for the core of the genome had built contigs for each chromosome that were close enough to overlap the telomere-specific clones (Goffeau et al. 1996, 1997). For other projects, there were small-telomere-containing clones (Consortium 1998; Wood et al. 2002), but these did not map back onto the genomes as the core chromosome contigs did not extend far enough into the subtelomeres. The gaps for some projects such as *C. elegans* and *S. pombe* have been slowly filled with a great deal of effort. For genomes with more chromosomes and therefore more telomeres, such as humans, and those with complex repetitive structures of the subtelomeres, such as *P. falciparum* and *T. brucei*, the screening of libraries and subsequent gap filling would take many person-years of labour (see *T. brucei* for example (Becker et al. 2004; Berriman et al. 2005; Hertz-Fowler et al. 2008) and the human subtelomere project (Riethman 1997, 2008a, b; Riethman et al. 1989, 2001, 2004, 2005). For some genome projects, generalizable techniques have worked well such as in various fungi, where a fosmid library approach for enriching telomere-containing clones resulted in a large insert library that could be mapped back to the core genome (Dean et al. 2005; Farman 2011; Li et al. 2005; Rehmeier et al. 2006; Wu et al. 2009).

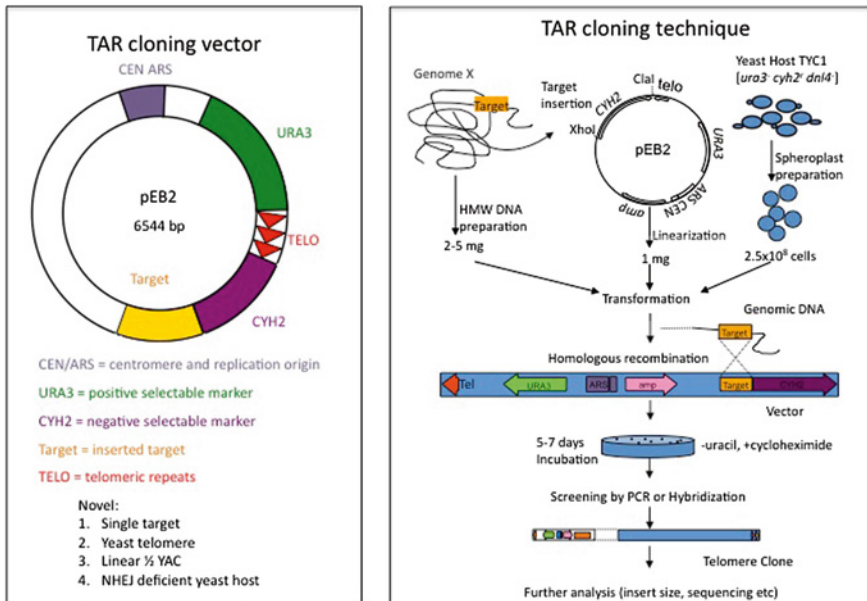
### 15.2.2 *Yeast to the Rescue I*

Yeast first came to the rescue by its ability to recognize telomere sequences from other organisms as a telomere (Szostak and Blackburn 1982) and its ability to tolerate large extra chromosomes of foreign genomes as YACs (Burke et al. 1987). In contrast to *E. coli*, AT-rich genomic fragments and long inverted repeats are more stable in yeast (Gardner et al. 2002; Glockner et al. 2002; Hayashi et al. 1993). The original YAC approach involved two telomeres with yeast markers, which were ligated onto the ends of large genomic fragments from another organism. Transformation of this into yeast usually resulted in a YAC with the markers at both ends, though occasionally YACs with only one added telomere came through and these had ‘captured’ a telomere from the other organism. A half-YAC approach to the human genome projects’ telomeres and subtelomeres then ensued (Riethman et al. 1989), though it has taken over 15 years to almost complete a set of human telomere and subtelomere clones as YACs (Riethman 2008b; Riethman et al. 2001, 2004, 2005). This approach has worked for other organisms as well such as *Pneumocystis carinii*, which was unculturable at the time, but was labour intensive and not efficient (Underwood et al. 1996). One problem with this approach was that many of the YACs were chimeras, containing fragments from more than one genomic location (Larionov et al. 1996a, b; Underwood et al. 1996).

### ***15.2.3 Yeast to the Rescue II: Transformation-Associated Recombination (TAR) Cloning***

Yeast came to the rescue a second time through the combination of several techniques and approaches into an elegant and generalizable method for cloning genes as well as large chromosomal fragments of up to 300 kb as YACs, called transformation-associated recombination (TAR) cloning (Larionov et al. 1996a). The technique is based on simultaneous transformation of yeast spheroplasts with genomic DNA and a TAR vector containing gene or sequence-specific targeting sequences (hooks) of minimally 60 bp length. Homologous recombination in the yeast cell between targeting sequences in the vector and the complementary, chromosomal DNA sequence, captures the chromosomal region between the targeting hooks as circular YAC molecules (Kouprina and Larionov 2006; Larionov et al. 1996a; Noskov et al. 2001, 2003). These are faithfully replicated and segregated in the yeast host alongside its usual unaltered set of chromosomes (Kouprina and Larionov 2006). Positive recombinants are selected for further analysis using PCR or hybridization-based screening methods. Yeast has several properties that have made this possible. The high rates of homologous recombination and the use of positive and negative selectable markers (*HIS3*, *URA3*) produce positive YAC recombinants at high rates (up to 40 %) and suppress negative background caused by vector recircularization from non-homologous end-joining (Kouprina and Larionov 2006; Noskov et al. 2002). The transformation efficiency of yeast is 100 times higher than *E. coli* and some human DNA sequences, including coding DNA, that were unstable in *E. coli*, and were therefore entirely missed, are stable in yeast (Kouprina et al. 2003). This combination led to the development of very efficient gap repair of plasmids transformed into yeast by recombination with homologous sequences in the genome (Ma et al. 1987). The amount of homology required could be very small and diverse, less than 60 base pairs and up to 15 % sequence divergence is tolerated. Although chromosomal recombination is greatly reduced in the presence of mismatches in the interacting DNA molecules, the recombination associated with transformation is tolerant of high levels of mismatches (up to 30 % divergence) (Larionov et al. 1994). These were combined to create the TAR cloning method which remarkably could result in large circular YACs using short Alu repeats as homologous targets in their vector (Larionov et al. 1996a). Numerous improvements over the years to increase efficiency have been made, including counter-selectable markers for enriching for recombinants, leaving the yeast origins out of the vector as the short consensus that functions in yeast can be found randomly in foreign genome sequences, development of specific sequence targets, vector improvements for movement into *E. coli* as BACs, etc. (Kouprina and Larionov 2006, 2008). The generation of YACs by TAR has several advantages over the original YAC method including no chimeras and the ability to target specific genomic locations.

## Principle of subtelomere trapping using TAR cloning



**Fig. 15.1** The principle of TAR cloning of subtelomeres. The trapping of subtelomeres by TAR is based on the use of a target sequence, homologous recombination and the fact that telomere repeats from most organisms function to seed new yeast telomeres, and the requirement of telomeres on both ends of a linear molecule in order to be maintained in yeast. As used to clone the subtelomeres of *T. brucei* (Becker et al. 2004), the vector contains all the necessary elements for replication and maintenance in yeast (centromere, one telomere, origin of replication (ARS element)), as well as a positive selectable marker (URA3), a counter-selectable marker (CYH2 conferring dominant sensitivity to cycloheximide) and the homologous target sequence. The yeast strain is deficient in *ura3*, has a recessive marker for cycloheximide resistance (*cyh2r*) and is deficient in the non-homologous end-joining specific ligase (*dnl4*). Co-transformation of the genomic DNA of interest with the linear vector into appropriately treated yeast cells results in colonies after 1 week on selective URA media. These are then ready for screening and further analysis

### 15.2.4 TAR Cloning of Subtelomeres

The existing TAR cloning method was modified to specifically capture telomeric and subtelomeric sequences and was first successfully used to isolate *T. brucei* subtelomeres (Becker et al. 2004). Firstly, a purpose-built basic vector was constructed in which there was a single targeting hook and a yeast telomere. In addition, the vector contains the yeast selectable marker URA3, a counter-selectable marker CYH2, a yeast centromere and an origin of replication (ARS). As shown in Fig. 15.1, a successful targeted recombination event traps a telomere from the hook to the end of the chromosome of interest, as telomeres from virtually any

organism function to seed new yeast telomeres. Selection for URA3 and against CYH2 enriches for the desired recombinants. Secondly, the deletion of the non-homologous end-joining specific ligase gene (DNL4) created a highly efficient yeast strain (ura3-, leu2-, dnl4-, cyh2-recessive resistance to cycloheximide), resulting in a threefold increase in the frequency of subtelomere clones over ligase-positive yeast (Becker and Louis unpublished results). The vector acts as a telomere trap whereby the yeast telomere repeats serve as a telomere on one end and the telomere of the genome of interest is captured using subtelomeric DNA sequence as targeting hook, resulting in linear half YACs, with a vector derived telomere on one and captured telomere on the other end of the linear molecule. The entire procedure, as shown in Fig. 15.1, can be conducted for multiple samples simultaneously within 7 days and typically generates thousands of recombinants.

The targeting sequence can be subtelomeric specific, either shared between many subtelomeres or specific to a given chromosome end, or they can be generic repeated elements such as transposable elements, few genomes being without any. Using the shared promoter region for the blood stream form expression sites (BES) of *T. brucei*, most expression sites of several isolates of *Trypanosoma* have been isolated (Becker et al. 2004; Young et al. 2008) and subsequently sequenced (Hertz-Fowler et al. 2008). The use of more dispersed transposable element repeats has successfully been used on *T. brucei*, *Brugia Malaya* and the planarian *Schmidtea mediterranea* (Becker and Louis, unpublished). Chromosome-end-specific targets have been used to clone individual telomeres in *S. pombe* (Becker and Louis, unpublished) as well as *T. brucei* (see databases (Aslett et al. 2010; Logan-Klumpler et al. 2012)). The use of two hooks flanking subtelomeric genes of the VAR gene family of *P. falciparum* successfully generated a library of the diverse flanking subtelomeric genes virulence factor from a novel isolate and could be used to assess diversity in endemic areas of infection (Gaida et al. 2011).

The frequency of subtelomere-positive clones in these TAR clone libraries was up to 30 % which is up to 100-fold higher than in many standard libraries. This demonstrates that subtelomeres and telomeres can be cloned from any genome even when little information is available using any the following 3 basic strategies: (1) For genomes that contain a known conserved subtelomeric sequence of at least 60 bp, multiple subtelomeres can be cloned simultaneously using a single TAR vector containing this element as targeting hook. This has been used to isolate subtelomeric blood stream form expression sites (BES) from *T. brucei* and *Trypanosoma brucei gambiense* using a conserved promoter element found at all BESs as targeting hook. These subtelomere libraries contained BES-positive clones at a frequency of up to 26 % with clone sizes ranging from 20 to 150 kb. This provided valuable insight into the architecture of BESs and aspects of their use in host adaptation and immune evasion (Becker et al. 2004; Hertz-Fowler et al. 2008; Young et al. 2008). (2) The cloning of specific telomeres using a unique sequence as targeting hook is applicable if telomere-proximal sequences are available that can be used to construct TAR vectors containing chromosome-end-specific targets. This method was successfully used to isolate up to 230 kb of subtelomeric regions of 14 missing chromosome ends of *T. brucei* for the genome project



using the end-most unique sequences of chromosome-specific contigs (see genome project databases (Aslett et al. 2010; Logan-Klumpler et al. 2012)). Here, the frequency of positive clones is less, 5.4 %, but still significantly more than standard libraries. (3) The cloning of multiple telomeres using dispersed repeated sequences or mobile genetic elements as targets is a useful strategy to clone subtelomeres in the absence of subtelomeric sequence information. This was successful in cloning subtelomeres from *T. brucei* using the 197-bp-RIME-A element and from *Brugia malayii* using the highly frequent HhaI element (Becker and Louis unpublished).

## 15.3 Bottlenecks

Despite the development of new cloning strategies including TAR cloning to generate subtelomere libraries, there are still bottlenecks in analyzing these clones, which prevent a wider use and affordable high-throughput approaches. Firstly, the purification of subtelomeric clones away from the yeast host genome is a time-consuming process, which involves several rounds of separation and isolation by pulsed-field gel electrophoresis with poor yields of enriched DNA. This has turned out to be very inefficient, for example taking 4 years to sequence a small set of the BES clones (Becker et al. 2004; Hertz-Fowler et al. 2008). The telomere-specific TAR clones for the genome project took longer and are still being analysed (Aslett et al. 2010; Logan-Klumpler et al. 2012). Secondly, the assembly of subtelomeres is still difficult due to their mosaic and repetitive nature. Even with subtelomeric sequences isolated away from others, the assembly of these regions has proven difficult due to the internal repetitive regions. Current assemblers cannot handle this complex repetitive structure. Developing such methodologies is particularly pertinent, considering the increasing reliance on genomic data generated using second-generation sequencing platforms with diminishing resources dedicated to targeted finishing, which is traditionally the only realistic way of tackling assemblies of subtelomeric regions. This is even more of a problem when many individuals are to be sequenced such as with the 1,000 human genome project (Kuehn 2008) or the population genomics of yeast project (Liti et al. 2009) with the desire to map genetic causes of phenotypic variation.

### 15.3.1 Possible Solutions

Some of the technical problems will be solved soon or could have solutions in existing technologies. Underrepresentation of telomeric and subtelomeric sequences is likely no longer a problem as second-generation sequencing has no cloning steps in *E. coli*. The short reads of most approaches exacerbates the assembly problem; however, the isolation of individual subtelomeres away from the rest of the subtelomeres of an organism helps with some of the assembly

issues. There is still the issue of purification of the clones from the yeast host. There are a few potential solutions to this:

1. Sequence the whole yeast genome along with the YAC without any purification. Although the YAC DNA will represent only 2 % of the DNA to be sequenced, current costs and coverage with short second-generation reads make this an affordable and relatively fast solution despite the obvious inefficiency.
2. Oligo-affinity enrichment: Each subtelomere containing YAC has the unique cloning vector at one end. Hybridization with a high-affinity oligo, using custom-made peptide nucleic acids (Chandler et al. 2000) for example, can be used to enrich for the vector and its attached subtelomeric DNA. This may not efficiently enrich the sequences far from the vector on long clones.
3. Use the old standard of CsCl gradients for genome with a different GC content than yeast.
4. Subtract the yeast genome DNA by targeted affinity capture leaving the subtelomeric YAC in solution.

The problem of assembly of repetitive regions is a generic one and in projects where the subtelomeres are not individually cloned remains a big problem, particularly with shorter reads of the current sequencing technologies. A possible solution will come through third-generation sequencing of single long molecules, which may span the shared homology regions of telomeres.

## 15.4 The Future

The study and analysis of subtelomeres has come a long way, and there are now reasonably efficient approaches towards completing individual genomes. One of the remaining big challenges will be population genomics, genome-wide association studies and quantitative genetics involving the subtelomeres. For this, there will have to be more rapid and efficient high-throughput means to obtaining the ends of chromosomes for many individuals. Yeast is a case in point where advances are being made in determining the underlying genetic cause of phenotypic variation. In genetic crosses between 4 different strains of yeast, used to map the genes responsible for a number of phenotypes, 25 % of the genes responsible for any given phenotype mapped beyond the last known segregating marker (Cubillos et al. 2011). This is a significant lack of understanding of quantitative traits and is likely to hold in other organisms such as humans and the study of disease-causing loci through genome-wide association studies. In yeast, the missing subtelomeric sequences represent about 8 % of the genome, indicating an enrichment of genes of interest in this unknown genomic region (Liti and Louis 2012). Even if there is no enrichment for such loci in human studies, there must be a great deal of genetic information on polygenic traits and disease in humans missing as they are in the subtelomeric regions. Not only are the subtelomeres interesting in their own right, having exciting biology as seen in the previous chapters, there is more biology to learn that we are not even aware of yet.

## References

- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science*, *287*, 2185–2195.
- Aslett, M., Aurrecochea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., et al. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Research*, *38*, D457–D462.
- Becker, M., Aitchison, N., Byles, E., Wickstead, B., Louis, E., & Rudenko, G. (2004). Isolation of the repertoire of VSG expression site containing telomeres of *Trypanosoma brucei* 427 using transformation-associated recombination in yeast. *Genome Research*, *14*, 2319–2329.
- Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., et al. (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science*, *309*, 416–422.
- Berriman, M., Hall, N., Shearer, K., Bringaud, F., Tiwari, B., Isobe, T., et al. (2002). The architecture of variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, *122*, 131–140.
- Burke, D. T., Carle, G. F., & Olson, M. V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*, *236*, 806–812.
- Celniker, S. E., Wheeler, D. A., Kronmiller, B., Carlson, J. W., Halpern, A., Patel, S., et al. (2002). Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biology*, *3*, RESEARCH0079.
- Chandler, D. P., Stults, J. R., Anderson, K. K., Cebula, S., Schuck, B. L., & Brockman, F. J. (2000). Affinity capture and recovery of DNA at femtomolar concentrations with peptide nucleic acid probes. *Analytical Biochemistry*, *283*, 241–249.
- Consortium, C. E. G. (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, *282*, 2012–2018.
- Costa, V., Casamassimi, A., Roberto, R., Gianfrancesco, F., Matarazzo, M. R., D'Urso, M., et al. (2009). DDX11L: A novel transcript family emerging from human subtelomeric regions. *BMC Genomics*, *10*, 250.
- Cubillos, F. A., Billi, E., Zorgo, E., Parts, L., Fargier, P., Omholt, S., et al. (2011). Assessing the complex architecture of polygenic traits in diverged yeast populations. *Molecular Ecology*, *20*, 1401–1413.
- Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., et al. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*, *434*, 980–986.
- Farman, M. L. (2011). Targeted cloning of fungal telomeres. *Methods in Molecular Biology*, *722*, 11–31.
- Farman, M. L., & Leong, S. A. (1995). Genetic and physical mapping of telomeres in the rice blast fungus, *Magnaporthe grisea*. *Genetics*, *140*, 479–492.
- Gaida, A., Becker, M. M., Schmid, C. D., Buhlmann, T., Louis, E. J., & Beck, H. P. (2011). Cloning of the repertoire of individual *Plasmodium falciparum* var genes using transformation associated recombination (TAR). *PLoS ONE*, *6*, e17782.
- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, *419*, 498–511.
- Glockner, G., Eichinger, L., Szafranski, K., Pachebat, J. A., Bankier, A. T., Dear, P. H., et al. (2002). Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. *Nature*, *418*, 79–85.
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., et al. (1996). Life with 6000 genes. *Science*, *274*(546), 547–563.
- Goffeau, A., et al. (1997). The yeast genome directory. *Nature*, *387*, 1–105.
- Hayashi, Y., Heard, E., & Fried, M. (1993). A large inverted duplicated DNA region associated with an amplified oncogene is stably maintained in a YAC. *Human Molecular Genetics*, *2*, 133–138.

- Hertz-Fowler, C., Figueiredo, L. M., Quail, M. A., Becker, M., Jackson, A., Bason, N., et al. (2008). Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PLoS ONE*, *3*, e3527.
- Kouprina, N., & Larionov, V. (2006). TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nature Reviews Genetics*, *7*, 805–812.
- Kouprina, N., & Larionov, V. (2008). Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nature Protocols*, *3*, 371–377.
- Kouprina, N., Leem, S. H., Solomon, G., Ly, A., Koriabine, M., Otstot, J., et al. (2003). Segments missing from the draft human genome sequence can be isolated by transformation-associated recombination cloning in yeast. *EMBO Reports*, *4*, 257–262.
- Kuehn, B. M. (2008). 1000 Genomes Project promises closer look at variation in human genome. *JAMA*, *300*, 2715.
- Larionov, V., Kouprina, N., Eldarov, M., Perkins, E., Porter, G., & Resnick, M. A. (1994). Transformation-associated recombination between diverged and homologous DNA repeats is induced by strand breaks. *Yeast*, *10*, 93–104.
- Larionov, V., Kouprina, N., Graves, J., Chen, X. N., Korenberg, J. R., & Resnick, M. A. (1996a). Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proceedings of the National Academy of Sciences*, *93*, 491–496.
- Larionov, V., Kouprina, N., Graves, J., & Resnick, M. A. (1996b). Highly selective isolation of human DNAs from rodent-human hybrid cells as circular yeast artificial chromosomes by transformation-associated recombination cloning. *Proceedings of the National Academy of Sciences*, *93*, 13925–13930.
- Li, W., Rehmeier, C. J., Staben, C., & Farman, M. L. (2005). TERMINUS—telomeric end-read mining in unassembled sequences. *Bioinformatics*, *21*, 1695–1698.
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., et al. (2009). Population genomics of domestic and wild yeasts. *Nature*, *458*, 337–341.
- Liti, G., & Louis, E. J. (2012). Advances in quantitative trait analysis in yeast. *PLoS Genetics*, *8*, e1002912.
- Liti, G., Nguyen Ba, A. N., Blythe, M., Muller, C. A., Bergstrom, A., Cubillos, F. A., et al. (2013). High quality de novo sequencing and assembly of the *Saccharomyces arboricolus* genome. *BMC Genomics*, *14*, 69.
- Logan-Klumpler, F. J., De Silva, N., Boehme, U., Rogers, M. B., Velarde, G., McQuillan, J. A., et al. (2012). GeneDB—an annotation database for pathogens. *Nucleic Acids Research*, *40*, D98–108.
- Louis, E. J. (1994). Corrected sequence for the right telomere of *Saccharomyces cerevisiae* chromosome III. *Yeast*, *10*, 271–274.
- Louis, E. J. (1995) Use of expand PCR system to complete the telomeres for the yeast genome project. *Biochemica*, *3*, 25–26.
- Louis, E. J., & Borts, R. H. (1995). A complete set of marked telomeres in *Saccharomyces cerevisiae* for physical mapping and cloning. *Genetics*, *139*, 125–136.
- Ma, H., Kunes, S., Schatz, P. J., & Botstein, D. (1987). Plasmid construction by homologous recombination in yeast. *Gene*, *58*, 201–216.
- Mefford, H. C., & Trask, B. J. (2002). The complex structure and dynamic evolution of human subtelomeres. *Nature Reviews Genetics*, *3*, 91–102.
- Noskov, V., Kouprina, N., Leem, S. H., Koriabine, M., Barrett, J. C., & Larionov, V. (2002). A genetic system for direct selection of gene-positive clones during recombinational cloning in yeast. *Nucleic Acids Research*, *30*, E8.
- Noskov, V. N., Koriabine, M., Solomon, G., Randolph, M., Barrett, J. C., Leem, S. H., et al. (2001). Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning. *Nucleic Acids Research*, *29*, E32.
- Noskov, V. N., Kouprina, N., Leem, S. H., Ouspenski, I., Barrett, J. C., & Larionov, V. (2003). A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast. *BMC Genomics*, *4*, 16.

- Oliver, S. G., Van der Aart, Q. J. M., Agostoni-Carbone, M. L., Aigle, M., Alberghina, L., Alexandraki, D., et al. (1992). The complete DNA sequence of yeast chromosome III. *Nature*, *357*, 38–46.
- Rehmeyer, C., Li, W., Kusaba, M., Kim, Y. S., Brown, D., Staben, C., et al. (2006). Organization of chromosome ends in the rice blast fungus, *Magnaporthe oryzae*. *Nucleic Acids Research*, *34*, 4685–4701.
- Riethman, H. (1997). Closing in on telomeric closure. *Genome Research*, *7*, 853–855.
- Riethman, H. (2008a). Human subtelomeric copy number variations. *Cytogenet Genome Res*, *123*, 244–252.
- Riethman, H. (2008b). Human telomere structure and biology. *Annual Review of Genomics and Human Genetics*, *9*, 1–19.
- Riethman, H., Ambrosini, A., Castaneda, C., Finklestein, J., Hu, X. L., Mudunuri, U., et al. (2004). Mapping and initial analysis of human subtelomeric sequence assemblies. *Genome Research*, *14*, 18–28.
- Riethman, H., Ambrosini, A., & Paul, S. (2005). Human subtelomere structure and variation. *Chromosome Research*, *13*, 505–515.
- Riethman, H. C., Moyzis, R. K., Meyne, J., Burke, D. T., & Olson, M. V. (1989). Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast-artificial-chromosome vector. *Proceedings of the National Academy of Sciences*, *86*, 6240–6244.
- Riethman, H. C., Xiang, Z., Paul, S., Morse, E., Hu, X. L., Flint, J., et al. (2001). Integration of telomere sequences with the draft human genome sequence. *Nature*, *409*, 948–951.
- Schwartz, S. L., & Farman, M. L. (2010). Systematic overrepresentation of DNA termini and underrepresentation of subterminal regions among sequencing templates prepared from hydrodynamically sheared linear DNA molecules. *BMC Genomics*, *11*, 87.
- Szostak, J. W., & Blackburn, E. H. (1982). Cloning yeast telomeres on linear plasmid vectors. *Cell*, *29*, 245–255.
- Underwood, A. P., Louis, E. J., Borts, R. H., Stringer, J. R., & Wakefield, A. E. (1996). *Pneumocystis carinii* telomere repeats are composed of TTAGGG and the subtelomeric sequence contains a gene encoding the major surface glycoprotein. *Molecular Microbiology*, *19*, 273–281.
- Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., et al. (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature*, *415*, 871–880.
- Wu, C., Kim, Y. S., Smith, K. M., Li, W., Hood, H. M., Staben, C., et al. (2009). Characterization of chromosome ends in the filamentous fungus *Neurospora crassa*. *Genetics*, *181*, 1129–1145.
- Young, R., Taylor, J. E., Kurioka, A., Becker, M., Louis, E. J., & Rudenko, G. (2008). Isolation and analysis of the genetic diversity of repertoires of VSG expression site containing telomeres from *Trypanosoma brucei* gambiense, *T. b. brucei* and *T. equiperdum*. *BMC Genomics*, *9*, 385.