OPEN ACCESS IOURNAL

BIOINTERFACE RESEARCH IN APPLIED CHEMISTRY

REVIEW ARTICLE www.BiointerfaceResearch.com ISSN 2069-5837

Volume 3, Issue 2, 2013, 523-532

Received: 10.02.2013 / Accepted: 25.03.2013 / Published on-line: 15.04.2013

Genome mapping of medically important yeasts

Ortansa Csutak [1]*, Tatiana Vassu[1]

ABSTRACT

Physical and genetic maps are important steps in the process of yeast genome exploration. One of the first strategies of physical mapping comprised a combination of whole genome shotgun and restriction analysis of large fragments cloned in BAC/YAC libraries. During last years, new approaches have emerged, such as screening of BAC libraries using PCR assays, BAC pooling strategies or *in silico* anchoring of BAC clones. Since the first physical map of *Saccharomyces cerevisiae* published in 1991 and the international project of its genome sequencing accomplished in 1996, the yeast research community developed a series of interconnected programs and databases as important tools in genome evolution understanding, genome annotation and comparative genomics. Advanced genome research and mapping revealed new insights on the genetic basis of pathogenesis related processes of many yeast species of medical importance belonging mainly to *Candida* and *Schizosaccharomyces* genera.

Keywords: *yeasts; physical maps; genome annotation; Candida; Schizosaccharomyces.*

1. INTRODUCTION

Microorganisms are an excellent basis for genomic analysis due mainly to their relative simple genome when compared to more complex, multicellular organisms. Physical maps of chromosomes or genomes represent important steps in genome exploration. Fiers *at al.* [1] were the first to publish the complete nucleotide sequence of bacteriophage MS2 RNA, followed by Sanger *et al.* who described the full nucleotide sequence of bacteriophage phi X174 DNA [2], and later, in 1981, by Ineichen *et al.* who determined the sequence of the phage lambda genome [3]. During the 90's the partial or entire sequences of many genomes were described: *Marchantia polymorpha* [4], cytomegalovirus [5, 6], *Vaccinia* [7], *Haemophilus influenzae* [8] and *Methanococcus jannaschii* [9]. The physical map of *Saccharomyces cerevisiae* was described by Link and Olson [10] while genome sequencing began in 1989 at Université Catholique de Louvain-La-Neuve under the coordination of André Goffeau as part of an EU study. The project lead to the development of *Yeast Genome Consortium* with the involvment of more than 90 laboratories accross Europe, United States and Japan. The entire *S. cerevisiae* genome sequence was accomplished by Goffeau *et al.* [11] and the genomes of various hemiascomycetous yeast species were further investigated during the 2000's. The main approaches used for physical mapping were, for a long period of time, based on computerised contig assembly from small fragments obtained by whole genome shotgun or on restriction analysis of large fragments cloned in BAC (Bacterial Artificial Chromosomes), PAC (P1-derived Artificial Chromosomes) or YAC (Yeast Artificial Chromosomes) libraries and ordered in contigs. The *hybrid* maps comprising a combination of contig assembly from fragments obtained by whole genome shotgun and restriction analysis of large genome fragments cloned in BAC/YAC

[1] Department of Genetics, Faculty of Biology, University of Bucharest, 1-3 Aleea Portocalelor, 060101-Bucharest, Romania

*Corresponding author e-mail address: ***cs_ortansa@yahoo.fr***

libraries, have good continuity and are strongly supported by clone collections that cover long ranges of mapped regions. During last years, many new approaches have been developed in order to investigate the genome structure of various yeast species, such as screening of BAC libraries using PCR assays, BAC pooling strategies, *in silico* anchoring of BAC clones and a wide range of analyses using the Illumina systems.

In the present article we present the main aspects of construction of physical maps, the relation between physical and genetic maps and a general updated information on genome structure and assembly for some the most studied yeast species of medical importance.

2. PHYSICAL MAPS

2.1. Hybrid physical maps based on BAC / YAC libraries. The first step in construction of *hybrid* physical maps consists in fragmenting the genomic DNA using, in paralell, whole genome shotgun and endonuclease digestion. The whole genome shotgun approach consists in breaking many copies of the entire genome into small fragments of 4 - 5 kbp which are cloned in plasmid or phagemid vectors, sequenced and assembled in contigs that cover each chromosome. These fragments can be used as markers in mapping, are called STS (sequenced tagged sites) and can be represented by random fragments of DNA (RST) or they can orriginate from coding (expressed) sequences (EST).

The same genome is digested with restriction endonucleases. The large fragments obtained are separated by PFGE (Pulsed Field Gradient Gel Electrophoresis) [12-17], purified and cloned in BAC/YAC vectors [18, 19] to obtain libraries usually stored on 96 wells microplates. Since several copies of the genome are subject to endonuclease restriction, the clones cover *3X* or *10X* genome. There are more pro arguments for using BAC vectors in clone libraries and map construction when compared to YACs, such as higher transformation efficiency in *E. coli* cells *vs*. yeast cells used for YACs, the stable maintenance of 300 kbp fragments for over 100 generations, a low risk of recombination and easier manipulation due to its supercoiled circular covalent structure (YACs are used as linear vectors) [20].

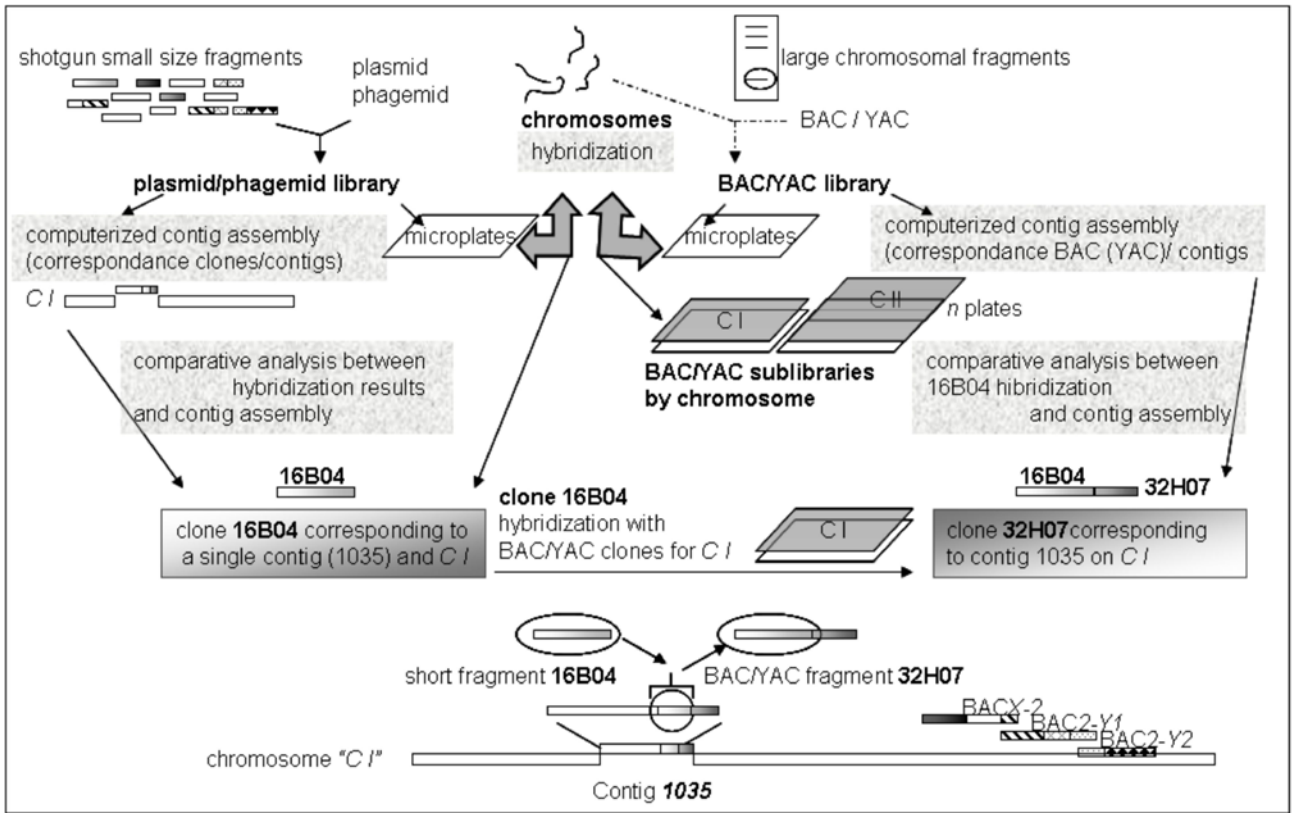

Figure 1: Construction of *hybrid* physical maps

Once the BAC/YAC library obtained, the chromosomal fragments are end-sequenced and assembled in contigs using specific software (*e.g.* PHRAP). The correlation between clones and each chromosome is obtained by hybridizing each microplate from the library with each chromosome isolated by PFGE. The colonies thus organized in sublibraries have to provide a good rate of coverage of the chromosomes, with the exception of the chromosome bearing rDNA genes organized in a cluster of large number of copies.

The small fragments (STS) are amplified by PCR and used for screening of BAC/YAC sublibraries. The colinearity of contigs and screened BAC/YAC clones allows the selection of a reduced number of chromosomal fragments which are assembled with minimal overlapping in order to ensure a good coverage of each chromosome (Fig. 1). (STS from clone 16B04 (where 16 is the number of the microplate, B04 is the position of the clone on the microplate) hybridized only with chromosome "*C I*" and was determined as belonging to contig 1035. Contig 1035 is found only on chromosome "*C I*" and comprises fragments from BAC/YAC clones 19A09 and 32H07. From these, only 32H07 is chosen for fingerprinting since it offers a minimal overlapping with the next BAC/YAC fragments along chromosome "*C I*")

2.2. BAC fingerprinting and BAC end-sequencing. The correct assembly and colinearity of contigs are verified using BAC/YAC fingerprints. On this purpose, the DNA fragments inserted in the BAC/YAC clones selected for each chromosome are digested with a single endonuclease, the patterns obtained are run through agarose gel electrophoresis, analysed and assembled using specific algoritms and softwares from Applied Biosystems and Life Technologies or the FPC software (FingerPrinted Contigs) [21]. In fact, the fingerprinting maps present the order and distance between enzyme cleavage sites. The main advantages of agarose gel fingerprinting reside in its simplicity and the possibility of observing all the restriction fragments obtained from a clone, the integrity of the overlap among clones being easy verified and estimated [22]. Moreover, the method allows the calculation of the size of each cloned fragment. Information on BACs fingerprinting was extensively used for constructing maps for human, yeasts, apple, fungi, mice or bacteria genomes [23-26]. The YAC based maps were used in rice and *Arabidopsis thaliana* research [27-29].

End-sequencing of large insert clones selected for each chromosome from whole genome libraries (BAC end-sequencing or BES) represents a reliable approach for genome mapping when combined with fingerprinting [30]. It consists in bidirectional sequencing of DNA chromosomal fragments from the clones with the help of universal primers for cloning sites from vector and is useful in ordering BAC clones along the chromosomes. BES information combined with identification of associated polymorphic markers is used in comparative genome analysis between related species [31]. The gaps between subsequent clones are usually covered by primer-walking. The obtained sequences are BLASTed against already known data and are classified as coding, non-coding or pseudo-genes.

2.3. New generation of mapping strategies in yeasts. Over the past years, new approaches have been developed in order to obtain more rapid and complete information on genome structure and functions. Many yeast mapping projects are based on whole genome sequencing which determines the total DNA sequence of chromosomal and mitochondrial DNA from the cell.

Although screening of BAC libraries by PCR assays is one of the techniques widley used in early mapping projects, the technique has been improved by development of new *in silico* analysis algorithms. Computer programs such as ELEPHANT (electronic physical mapping anchoring tool) are able to identify an accurate relationships between markers and BAC clones arranged in pools, reducing the number of PCR assays to be performed [32]. The BAC pooling strategies reduce the number of clones used in subsequent reactions by concentrating the entire number of clones from a

library into pools of overlaping clones. The most common strategies are two-dimensional (2D), three-dimesional (3D) and six-dimensional (6D) pooling [33].

Another approach is *in silico* anchoring of BAC clones to specific markers based on the sequence similarity. Therefore, knowledge from paired end sequencing of BAC clones is a benefit for anchoring the resulting physical contig map to the genetic map [34].

The Illumina systems are used in SNP discovery, BAC pooling [34], whole-genome, *de novo* and amplicon sequencing, target resequencing, identification of copy number variations and chromosomal rearrangement.

3. GENOME ANNOTATION. PHYSICAL MAPS - GENETIC MAPS

Genome annotation consists in identification of gene localisation and structure, determination of ORFs and regulatory regions (structural annotation) followed by description of their functions - structural, metabolic, regulatory (functional annotation). These processes are much elaborated and time consuming. The Gene Ontology (GO) consortium founded in 2000 offers a reliable and accurate solution to gene annotation. The GO is represented by a research community comprising participants working on a wide range of organisms: yeasts *Schizosaccharomyces pombe* (PomBase), bacteria *Escherichia coli* (PortEco), plant *Arabidopsis thaliana* (TAIR) or nematode *Caenorhabditis elegans* (WormBase) [35]. The GO allows a wide comparative analysis based on structured vocabularies or ontologies containing information obtained from genome sequencing or proteome analysis projects of various species, regarding gene products in terms of their molecular functions, the biological processes in which they are involved and the associated cellular components. A powerful tool developed by the GO Consortium is AmiGO used for browsing the Gene Ontology database for one or more gene products and its/their annotations.

Physical maps provide the basis for studying genome organization. Once the genes are cloned and described, the physical linkage can be determined and also a correlation can be established between physical and genetic distances. The genetic map offers the representations of genes order on chromosomes and the distance between them expressed in centimorgans, *i.e.* percentage of recombination between loci from the same chromosome. Although genes should be placed in the same order on the chromosome in physical and genetic maps, the distance between them might not be the same in both maps due to the variation of recombination frequencies which can be augmented in some chromosomal regions compared to others.

Physical and genetic maps are prone to error affecting the ability to establish an accurate order of the genes or other landmarks used as markers. In the case of physical maps, the errors are most probably derived from an incorrect assembly of clones or identification of marker positions on chromosomes [36], while for the genetic maps they might occur due to lack of enough information concerning meiosis recombination and to the presence of strain polymorphisms caused by subtelomeric repetitive sequences and Ty elements. Anyhow, it seems that inaccuracies between the two types of maps have multiple causes related to defaults in assembly of clones located near gaps that appear during analysis of physical-sequence data, to the presence of events like inversions and translocations and to difficulties in estimating recombination frequencies between genes/markers that are closely linked to each other. The knowledge acquired during development of yeast genome projects, shows that in order to obtain a proper ordering of genes on chromosomes, it is necessary to pursue independent physical mapping and sequence determination of a genome.

4. YEASTS OF MEDICAL IMPORTANCE

The augmenting number of yeast genomes completely sequenced allows a better exploration and understanding of gene order, structure and functions, being important tools in genome evolution and comparative genomics research of medically important yeasts [37].

The genus *Candida* comprises many species with pathogen potential. *C. albicans* is one of the most studied yeast species, mainly due to its high pathogenity for humans. The first physical map of haploid strain *C. albicans* 1006 was based on *Sfi* I macrorestriction of PFGE isolated chromosomes and revealed 34 restriction sites leading to chromosomal fragments ranging from 45 kbp to 2.2 Mb [38]. The fragments were assigned to the chromosomes using, in a first step, Southern blotting. Then the fragments neighboring those already assigned were subject to partial digestion. Finnaly, the isolated chromosomes were digested and hybridized with telomere-specific probes. Many translocations were identified, some of *Sfi* I sites being also interchromosomal recombination sites.

A new step in investigating the *C. albicans* genome was assembling the diploid genome sequence of heterozygous diploid strain *C. albicans* SC5314 [39]. The methods used were whole genome shotgun and a new developed PHRAP assembly method, the result being designated as Assembly 19. A number of 121 supercontigs were ordered on chromosomes with high coverage rates (82.7% for chromosome 6 to 95.9% for chromosome 1), with 26 supercontigs for chromosome 1 (the largest - 3165 kbp) and only 8 supercontigs for chromosome 7 (1020 kbp). The exception was chromosome R which contains the rDNA cluster, the arrangement of these genes being similar to that of *S. cerevisiae* plus a 2 kbp region of low complexity. The final assembly comprised data on protein encoding genes and region regulatory sequences polymorphisms important in pathogenesis. Thus, there have been described large gene families: the ALS (agglutinin-like sequence) family comprises 19 genes encoding for large cell-surface glycoproteins that are implicated in the process of adhesion to host surfaces and biofilm formation, there are 20 genes involved in iron transport, 21 secreted aspartyl proteinase (SAPs) and 22 secreted lipases genes. Also, it seems that *C. albicans* presents 23 genes that encode for a family of chloride channels resembling to some types expressed in human tissues. Moreover, studies on the haploid set revealed additional gene families (oligopeptide transport, estrogen-binding protein and acid sphingomyelinases) with roles in infection.

In 2007, haploid Assembly 21 contained 15.845 Mb DNA and showed that all the chromosomes bear a sub-telomeric repeat from chromosome 7, CARE-2 (*C. albicans* repetitive element 2) is present on all chromosomes except 2 and 7, while *TLO* (telomeric associated) genes probably involved in pathogenesis are missing only from one telomeric region on each chromosome 2 and 7 showing sequence similarity with three *C. dubliniensis* genes [40]. The analysis of SAP (secreted aspartyl proteinases) and LIP (lipase) gene families revealed a new model of dispersion: if a family which exists and expands on one chromosome is duplicated on another chromosome, it may follow the same process again.

In 2009, a comparative study of the genomes of the closely related species *C. dubliniensis* and *C. albicans*, revealed new insights on the genetic basis of pathogenesis processes in the two species [41]. The diploid genome of *C. dubliniensis* has 14.6 Mb with ten haploid and three diploid chromosomes. The 11X coverage of *C. dubliniensis* genome was sequenced by whole genome sequencing, assembled in 1110 contigs and the OrthoMCL algorithm was used for a comparative analysis with the *C. albicans* genome. The reduced pathogenicity of *C. dubliniensis* compared to *C. albicans* seems to have multiple causes. Thus, massive loss of genes was observed in *C. dubliniensis* genome. From the 115 pseudogenes found in *C. dubliniensis*, some are orthologs of the filamentous growth regulator genes from *C. albicans*. Also, the *ALS3* member of ALS gene family is absent in *C.*

dubliniensis. The *SAP5* and *SAP6* genes from SAP family, evolved only in *C. albicans* most probably by sequential segmental inversion, providing a better hyphal development and an increased virulence. A very important fact is the absence in *C. dubliniensis* of *HYR1* gene encoding a hyphal-induced cell-wall protein that confers resistance to killing by neutrophiles and to azole. All these data confirm the hypothesis that *C. dubliniensis* derived as a defective pathogen from a more virulent ancestor.

C. parapsilosis who is responsible for many *Candida* outbreaks, has been divided in three groups who became species: I – *C. parapsilosis*, II – *C. orthopsilosis* and III – *C. metapsilosis*, the last two being less virulent than *C. parapsilosis*. However, *C. orthopsilosis* seems to be the closest to *C. parapsilosis*, being able to form smaller biofilms and producing damage of human epithelial tissue. In order to identify and analyze the genes involved in pathogenicity, the 10X coverage of the genome of *C. orthopsilosis* was sequenced and assembled *de novo* as a 12.6 Mb consensus with 39 scaffolds [42]. A fosmid library was used to verify the data and to merge the scaffolds in 8 superscaffolds corresponding largely to the chromosomes. The data obtained using the Illumina technology were integrated and the gene prediction and annotation were done using the *Candida* Gene Order Browser database. The comparative analysis with the *C. parapsilosis* genome, supports the idea that the higher virulence of *C. parapsilosis* is due to expansion of some gene families and presence of genes not detected in *C. orthopsilosis*. Thus, whereas the HYR/IFF family has 17 members in *C. parapsilosis*, most of them are absent on similar chromosomal sites in *C. orthopsilosis*. The ALS family consists of five members in *C. parapsilosis* (one on chromosome 5 and four in tandem, on chromosome 4) and three in *C. orthopsilosis* (one on chromosome 5 and two on chromosome 4). An important feature of *C. orthopsilosis* and *C. parapsilosis* is the increased induced resistance to antifungal drugs compared to *C. albicans*. It seems that this is related to the efflux pumps gene, especially to FLU1 (fluconazole resistance) and MDR1 (multidrug resistance gene): there are 8 members of FLU1 group in *C. parapsilosis* and 6 in *C. orthopsilosis*, and an additional pair of Mdr1-like protein.

Candida (Torulopsis) glabrata is, after *C. albicans*, the most common microorganism isolated from fungal infections from human mucosal tissues and immunocompromised patients. The 8X coverage of genome of strain *C. glabrata* CBS138 was sequenced and mapped. It has 12.3 Mb and is organized in 13 chromosomes, 5283 coding genes and 207 tRNA genes [24]. For many years, no clear data were available regarding the sexual state and mating genes of *C. glabrata*, main features related to pathogenesis in *C. albicans*. In 2003, a genome survey sequencing study showed that *C. glabrata* CBS138 is closely related to *S. cerevisiae* in what concerns the genes involved in mating (13 orthologs of *S. cerevisiae* genes), meiosis and sporulation [43]. *C. glabrata* has orthologs of the genes from the MAP kinase pathway (*STE11*, *STE7*, *FUS3* or *STE12*), putative mating-type loci equivalent to MATalpha1 and MATalpha2, genes involved in karyogamy and loci involved in mating type interconversion [44]. Also, 19 genes with functions in meiosis and sporulation in *S. cerevisiae* have orthologs in *C. glabrata*, including *IME1* which activates early meiotic genes, *SPO1* for spindle body duplication and other meiosis specific genes (*SPO22*, *CSM1* and *CSM3*). These data suggest that *C. glabrata* might undergo a sexual life cycle just like *C. albicans*, and that *C. glabrata* and *C. albicans* have probably evolved independently as human pathogens from sexual reproducing ancestors.

The lack of hyphae formation or secreted proteolytic activity, indicates that *C. glabrata* might present new virulence mechanisms compared to *C. albicans*. Signature-tagged mutagenesis studies on 18350 *C. glabrata* Tn7 insertion mutants revealed that a number of 56 ORFSs and 20 intergenic regions are required for *C. glabrata* survival in THP-1 macropages [45]. These genes are involved in

chromatin organization (12%), DNA repair (9%), golgi vesicle transport (9%) and endocystosis (7%). *C. glabrata* cells adapt to the macrophage environment by remodeling their chromatin structure and functions, processes that involve genes *CgRSC3-A, CgRTT109, CgRTT107* and *CgSGC1*. Moreover, the genome-wide transcriptional profiling analysis showed that metabolic changes occur in internalized *C. glabrata* cells as response to glucose limitation, by down regulating the glycolysis and up-regulating glyoxylate and citrate cycles.

Further knowledge on genome assembly and gene annotation can be found on the *Candida* Genome Database [46], *Candida* Database from Broad Institute [47] and Génolevures [48].

Schizosaccharomyces pombe is one of the yeast species with features closely related to higher eukaryotes, including humans. Many genes and gene regulation mechanisms described in *S. pombe* proved to have similarity not only with those from *S. cerevisiae*, but also with those from *Drosophila melanogaster*, *Cernohabditis elegans*, *Arabidopsis thaliana* and *Homo sapiens*. The genome of *S. pombe* presented in 2002 was mapped using end-sequencing and restriction of 452 cosmid fragments, the gaps being covered using 13 long-range PCR products, 22 plasmid libraries and 15 BAC libraries [49]. The DNA fragments from the clones were sonnicated, sub-cloned in M13 or pUC18 vectors and random clones were sequenced and analysed. The final 13,8 Mb genome comprised 3 chromosomes of 5.7, 4.6 and 3.5 Mb and the mitochondrial DNA of 20 kbp. Gene prediction yield 4940 protein coding genes and 33 pseudogenes. Most usefull were the data obtained on the 50 *S. pombe* genes and gene products related to human diseases. From these, 23 can be related to human cancer and are represented by genes involved in checkpoints (cds1– human CHK2 protein-kinase), cell cycle (cdc2 – human CDK4, cdc17 - human immunodefficiency), DNA repair (swi9 or rad20 – Xeroderma pigmentosum, rad12 – Bloom syndrome) or genome stability (tel1 – Ataxia telangiectasia). The rest of 27 genes and gene products are implicated in metabolic, renal, cardiac and neurological disease.

Aslett and Wood [50] published a summary of gene ontology annotations for *S. pombe*. An expansion with 0.8% of the number of protein coding genes was achieved using a combined approach involving tandem mass-spectrometry and comparative genomics, and genome wide domain prediction [51].

Updated information on *S. pombe* genome sequencing, mapping and annotation can be found on PomBase [52].

5. CONCLUSIONS

Physical maps define the location of physical genes or identifiable DNA landmarks and the distance between them. A combined approach of whole genome shotgun and restriction-based fingerprinting of large chromosomal fragments can be succesfully followed for obtaining highly accurate results. The resulting *hybrid* maps have not only good continuity using a rather small number of overlapping chromosomal fragments, but are also supported by clone collections that cover long ranges of mapped regions. Genome annotation aims to determination of coding and regulatory regions (structural annotation) and description of their structural, metabolic and regulatory functions (functional annotation). The Gene Ontology (GO) consortium database facilitates the annotation process and a comparative analyses based on structured vocabularies that uses information from genome sequencing or proteome analysis of various species.

The new approaches developed over the past years, allow obtaining more accurate information on genome structure and functions and the genetic basis of pathogenesis related processes in various yeast species of medical importance belonging mainly to *Candida* and *Schizosaccharomyces* genera.

Finally, studies on combined physical and genetic maps showed that, in order to obtain a proper ordering of genes on chromosomes, it is necessary to pursue independent physical mapping and sequence determination of genome.

6. REFERENCES

[1] Fiers W., Contreras R., Duerinck F., Haegeman G., Iserentant D., Merregaert J. *et al.*, Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene, *Nature*, 260, 5551, 500-507, **1976.**

[2] Sanger F., Air G.M., Barrell B.G., Brown N.L., Coulson A.R., Fiddes C.A., Hutchinson C.A., Slocombe P.M., Smith M., Nucleotide sequence of bacteriophage phi X174 DNA, *Nature*, 265, 5596, 687-695, **1977.**

[3] Ineichen K., Shepherd J.C.W., Bickle T.A., The DNA sequence of the phage lambda genome between PL and the gene bet, *Nucleic Acids Res.*, 9, 18, 4639-4653, **1981**.

[4] Ohyama K., Chloroplast and mitochondrial genomes from a liverwort *Marchantia polymorpha* – gene organization and molecular function, *Biosci. Biotech. Biochem.*, 60, 1, 16-24, **1996**.

[5] Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R. *et al.*,Analysis of the protein-coding content of the sequence of the human cytomegalovirus strain AD169, *Curr. Top. Microbiol. Immunol.*, 154,125–170, **1990**.

[6] Dargan D.J., Jamieson F.E., MacLean J., Dolan A., Addison C., McGeoch D.J., The published DNA sequence of human cytomegalovirus strain AD169 lacks 929 base pairs affecting genes UL42 and UL43, *J. Virol*., 71, 12, 9833–9836, **1997**.

[7] Goebel S.J., Johnson G.P., Perkus M.E.., Davis S.W., Winslow J.P., Paoletti E., The complete DNA sequence of vaccinia virus, *Virology*, 179, 1, 247-266, 517-563, **1990**.

[8] Fleishmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R. *et al.*, Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd., *Science*, 269, 5223, 496-512, **1995**.

[9] Bult C.J., White O., Olsen G.J., Zhou L., Fleishmann R.D., Sutton G.G. *et al*., Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*, *Science*, 273, 5278, 1058-1073, **1996**.

[10] Link A.J., Olson M.V., Physical map of the *Saccharomyces cerevisiae* genome at 110-kilobase resolution, *Genetics*, 147, 681-698, **1991**.

[11] Goffeau A., Barrell B.G., Bussey H., Davis R.W., Dujon B., Feldmann H. *et al.*, Life with 6000 genes, *Science*, 274, 546-567, **1996**.

[12] Carle G.F., Olson M.V., Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis, *Nucleic Acids Res*., 12, 14, 5647-5664, **1984**.

[13] Schwartz D.C., Cantor C.R., Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis, *Cell*, 37, 67-75, **1984**.

[14] Carle G.F., Olson M.V., An electrophoretic karyotype for yeast, *Proc. Natl. Acad. Sci. USA*, 82, 3756-3760, **1985**.

[15] Carle G.F., Frank M., Olson M.V., Electrophoretic separation of large DNA molecules by periodic inversion of the electric field, *Science*, 232, 65-68, **1986**.

[16] Chu G., Pulsed field electrophoresis in contour-clamped homogenous fields for the resolution of DNA by size or topology, *Electrophoresis*, 10, 290-295, **1989**.

[17] Birren B., Lai E., *Pulsed field electrophoresis. A practical guide*, Academic Press Inc., San Diego, USA, **1993**.

[18] Shizuya H., Birren B., Kim U.J., Valeria M., Slepak T., Tachiiri Y., Simon M., Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector, *Proc. Natl. Acad. Sci. USA* , 89, 879-8797, **1992**.

[19] Bruschi C.V., Gjuracic K., *Yeast artificial chromosomes.* In: *Encyclopedia of life sciences.* Macmillan Publishers Ltd., Nature Publishing Group, London, UK, **2002**.

[20] She K., So you want to work with giants: The BAC vector, *BioTeach J.*, 1, 69-74, **2003**.

[21] Soderlund C., Humphray S., Dunham A., French L., Contig built with fingerprints, markers and FPC V4.7., *Genome Res.*, 10, 1772-1787, **2000**.

[22] Meyers B.C., Scalabrin S., Morgante M., Mapping and sequencing complex genomes: let's get physical, *Nature*, 5, 578-588, **2004**.

[23] Tao Q., Chang Y.-L., Wang J., Chen H., Islam-Faridi M.N., Scheuring C. *et al*., Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis, *Genetics*, 158, 1711-1724, **2001**.

[24] Dujon B, Sherman D, Fischer G. *et al,*. Genome evolution in yeasts, *Nature*, 430, 35-44, **2004**.

[25] Xu Z., van den Berg M.A., Scheuring C., Covaleda L., Lu H., Santos F.A. *et al.*, Genome physical mapping from large-insert clones by fingerprint analysis with capillary electrophoresis: a robust physical map of *Penicillium chrysogenum*, *Nucleic Acids Res*., 33, 5, 2-8, **2005**.

[26] Han Y., Gasic K., Marron B., Beever J.E., Korban S.S., A BAC-based physical map of the apple genome, *Genomics*, 89, 630-637, **2007**.

[27] Hauge B.M., Goodman H.M., Theory and application of YAC technology for genome research, *Probe*,1, http://www.nal.usda.gov/pgdic/Probe , **1991.**

[28] Kurata N., Umehara Y., Tanoue H., Sasaki T., Physical mapping of rice genome with YAC clones, *Plant Mol. Biol.*, 35, 101–113, **1997**.

[29] Sasaki T., The rice genome project in Japan, *Proc. Natl. Acad. Sci. USA*, 95, 2027–2028, **1998**.

[30] White P.S., Matise T.C., *Chapter 6. Genomic mapping and mapping databases.* In: *Bioinformatics: A practical guide to the analysis of genes and proteins. Second edition*, John Wiley & Sons Inc., USA, **2001**.

[31] Liu H., Jiang Y., Wang S., Ninwichian P., Somridhivej B., Xu P., Abernathy J., Kucuktas H., Liu Z., Comparative analysis of catfish BAC end sequences with the zebrafish genome, *BMC Genomics*, 10, 592-609, **2009**.

[32] Paux E., Legeai F., Guilhot N., Adam-Blondon A.-F., Alaux M., Salse J., Sourdille P., Leroy P., Feuillet C., Physical mapping in large genomes: accelerating anchoring of BAC contigs to genetic maps through in silico analysis, *Funct. Integr. Genomics*, 8, 29–32. **2008**.

[33] Vu G.T.H., Caligari P.D.S., Wilkinson M.J., A simple, high throughput method to locate single copy sequences from Bacterial Artificial Chromosome (BAC) libraries using High Resolution Melt analysis, *BMC Genomics*, 11, 301-308, **2010**.

[34] Ariyadasa R., Stein N., Advances in BAC-based physical mapping and map integration strategies in plants, *J. Biomed. Biotechnol*., ID 184854, 1-11, **2012**.

[35] Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry JM. *et al.* (The Gene Ontology Consortium). Gene Ontology: tool for the unification of biology, *Nat. Genet*., 25, 25-29, **2000**.

[36] DeWan A.T., Parrado A.R., Matise T.C., Leal S.M., The map problem: a comparison of genetic and sequence-based physical maps, *Am. J. Hum. Genet*., 70, 101–107, **2002**.

[37] Butler G., Rasmussen M.D., Lin M..F., Santos M.A.S., Sakthikumar S., Munro C.A., Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes, *Nature*, 459, 7247, 657–662, **2009**.

[38] Chu W.S., Magee B.B., Magee P.T., Construction of an SfiI macrorestriction map of the *Candida albicans* genome, *J. Bacteriol*., 175, 6637-6651, **1993**.

[39] Jones T., Federspiel N.A., Chibana H., Dungan J., Kalman S., Magee B.B. *et al.*, The diploid genome sequence of *Candida albicans*, *Proc. Natl. Acad. Sci. USA*, 101, 7329–7334, 2004.

[40] van het Hoog M., Rast T.J., Martchenko M., Grindle S., Dignard D., Hogues H. *et al.*, Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. *Genome Biol*., 8, R52, **2007**.

[41] Jackson A.P., Gamble J.A., Yeomans T., Moran G.P., Saunders D., Harris D. *et al.*, Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*, *Genome Res*., 19, 2231–2244, **2009**.

[42] Riccombeni A., Vidanes G., Proux-Wéra E., Wolfe K.H., Butler G., Sequence and analysis of the genome of the pathogenic yeast *Candida orthopsilosis*, *PLoS ONE*, 7, e35750, **2012**.

[43] Wong S., Fares M.A., Zimmermann W., Butler G., Wolfe K.H., Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast *Candida glabrata*, *Genome Biol*., 4, R10, **2003**.

[44] Bialková A., Šubik J., Biology of the pathogenic yeast *Candida glabrata*, *Folia Microbiol*., 51, 1, 3-20, **2006**.

[45] Rai M.N., Balusu S., Gorityala N., Dandu L., Kaur R., Functional genomic analysis of *Candida glabrata* - macrophage interaction: role of chromatin remodeling in virulence, *PLOS Pathog*., 8, e1002863, **2012**.

[46] Arnaud M.B., Inglis D.O., Skrzypek M.S., Binkley J., Shah P., Binkley G., Miyasato S.R., Simison M., Sherlock G., *Candida* Genome Database (http://www.candidagenome.org).

[47] *Candida* Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/).

[48] Sherman D.J., Martin T., Nikolski M., Cayla C., Souciet J.L., Durrens P., Génolevures Consortium. Génolevures: protein families and synteny among complete hemiascomycetous yeast proteomes and genomes, *Nucleic Acids Res*., 37 (Database issue), D550-4, **2009**.

[49] Wood V., Gwilliam R., Rajandream M.-A., Lyne M., Lyne R., Stewart A. *et al.*, The genome sequence of *Schizosaccharomyces pombe*, *Nature*, 415, 871-880, **2002**.

[50] Aslett M., Wood V., Gene Ontology annotation status of the fission yeast genome: preliminary coverage approaches 100%, *Yeast*, 23, 913–919, **2006**.

[51] Bitton D.A., Wood V., Scutt P.J., Grallert A., Yates T., Smith D.L., Hagan I.M., Miller C.J., Augmented annotation of the *Schizosaccharomyces pombe* genome reveals additional genes required for growth and viability, *Genetics*, 187, 1207–1217, **2011**.

[52] Wood V., Harris M.A., McDowall M.D., Rutherford K., Vaughan B.W., Staines D.M., Aslett M., Lock A., Bähler J., Kersey P.J., Oliver S.G., PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res.*, 40 (Database issue), D695-9, **2012**.