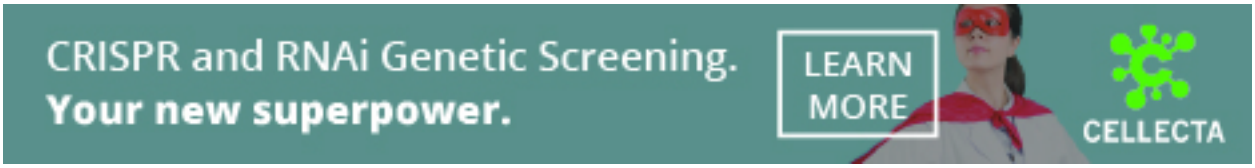
Assessment of colinearity between large cloned DNA fragments and genomic DNA.

P Verhasselt and G Volckaert

Genome Res. 1994 4: 129-132



Assessment of Colinearity between Large Cloned DNA Fragments and Genomic DNA

Peter Verhasselt and Guido Volckaert

Laboratory of Gene Technology, Catholic University Leuven, B-3001 Leuven, Belgium

Current strategies in genome sequencing projects ordinarily involve cloning of subgenomic DNA fragments into cosmid or phage λ vectors, arranging clones into a set of overlapping fragments, and the subsequent sequencing of overlapping clones individually. Any rearrangement occurring during cloning, amplification, or storage can lead to error-prone deductions and discredit the sequencing result. We propose a colinearity test based on comparative DNA amplifications by PCR with nested sets of primers on cloned versus genomic DNA, and spanning the entire sequenced region. This colinearity test is exemplified by analysis of a 32.5-kb region of *Saccharomyces cerevisiae* chromosome XIV.

Recent developments in DNA sequencing techniques have led to specific genome sequencing projects, often as an international endeavor (e.g., Yeast Genome Sequencing Programme of the European Community, Human Genome Organization). The genetic blueprint of *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, and *Caenorhabditis elegans* are under analysis, and projects to tackle even larger genomes, such as those of *Arabidopsis thaliana* or humans, have been set up or are under consideration.[1–5]

The Yeast Genome Sequencing Programme of the European Community follows a chromosome-by-chromosome approach and covers all or part of 9 of the 16 chromosomes of the budding yeast *S. cerevisiae*. The remainder of the genome is being sequenced by research teams in the United States, Canada, and Japan and at the Medical Research Council and Sanger Centre in Cambridge/Hinxten (UK). The usual strategy is to carve up individual chromosomes or the entire genome into cosmid or bacteriophage λ clones (recent additions are the use of bacteriophage P1[6] or YAC).[7] Chromosome-specific clones constituting a contiguous set of overlapping fragments are selected. The selected clones are sequenced individually and eventually the DNA sequence of the chromosome is reassembled from the sequences of those clones. This approach requires that the cosmid or phage clone sequences are exact copies of the chromosomal regions that they were derived from, a condition that apparently is not always met. Rearrangements usually are detected by restriction mapping and/or hybridization, after propagation of the clones. The resolution of these techniques is limited and highly dependent on the size of the restriction fragments. In this study we describe a method for the assessment of colinearity with a constant and high resolution. It is illustrated by the analysis of a 32.5-kb fragment from yeast chromosome XIV. Both the cosmid DNA and genomic DNA are scanned in parallel by PCR with nested sets of primers.

MATERIALS AND METHODS

Preparation of Genomic DNA

S. cerevisiae strain FY1679 was grown for 2 days at 28°C in 100 ml of YPD [1% (wt/vol) yeast extract, 2% (wt/vol) glucose, and 2% (wt/vol) peptone]. Cells were collected by centrifugation at 1000*g* for 10 min at 4°C, washed with 8 ml of water, and incubated for at least 1 hr at 30°C in 8 ml of lyticase solution [1 M sorbitol, 0.1 M K_2PO_4 (pH 7.5), 10 mM EDTA, 4 mg of lyticase (Sigma Chemical Co., St. Louis, MO), 0.1% (vol/vol) 2-mercaptoethanol]. Then, 8 ml of lysis buffer [2% (wt/vol) SDS; 50 mM Tris-HCl (pH 8.0); 10 mM EDTA] was added. After 10 min of moderate shaking at room temperature, 4 ml of 5 M NaCl was added. DNA was precipitated by incubation at 4°C for at least 2 hr, followed by centrifugation at 2000*g* for 30 min. The pellet was resuspended in 8 ml of TE buffer [10 mM Tris-HCl (pH 7.5), 1 mM EDTA], extracted with phenol/chloroform/isoamylalcohol (25:24:1), and precipitated with ethanol. The DNA pellet was redissolved in water to the desired concentration.

Preparation of Cosmid DNA

Cosmid 14-20 was obtained from P. Philippsen (Basel, Switzerland) and was isolated from a cosmid library (constructed by B. Dujon, Paris, France) of total DNA from yeast strain FY1679. This library consists of size-fractionated partial *Sau*3AI fragments cloned in the unique *Bam*HI site of the cosmid vector pWE15.[8] The *E. coli* clone containing cosmid 14-20 was grown overnight at 37°C with vigorous shaking in 200 ml of TB [17 mM KH_2PO_4, 72 mM K_2HPO_4, 1.2% (wt/vol) Bacto-tryptone, 2.4% (wt/vol) Bacto-yeast extract, 0.4% (vol/vol) glycerol]. Ampicillin was added to a final concentration of 25 μg/ml. Cosmid DNA was purified on a Qiagen-tip-500 ion ex-

4:129–132©1994 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/94 $5.00

change column (Diagen GmbH, Düsseldorf, Germany).

Amplification

PCR amplification[9] was performed by mixing 50 ng of cosmid DNA or 200 ng of total yeast DNA with 10 µl of buffer [100 mM Tris-HCl (pH 9.0), 500 mM KCl, 15 mM $MgCl_2$, 1% (vol/vol) Triton X-100, 0.1% (wt/vol) gelatine], 2 µl of each dNTP (10 mM), 5 µl of each primer (20 µM), and 0.2 units of SuperTaq (HT Biotechnology LTD, Cambridge, UK). This mixture was adjusted to 100 µl with sterile water and overlaid with 3 drops of mineral oil. DNA was amplified on a GeneAmp PCR System 9600 (Perkin-Elmer Cetus, Norwalk, CT) using 25 cycles of denaturation (30 sec at 95°C), primer annealing (30 sec at 45 °C), and DNA polymerization (5 min at 72°C). A 10-µl sample from each reaction was loaded onto a 0.8% agarose gel and electrophoresed at 80 V in Tris-acetate buffer for 2 hr. The gel was stained with ethidium bromide, and DNA bands were visualized under UV light and photographed.

Oligonucleotides

Oligonucleotides (15- and 16-mers) were synthesized by the β-cyanoethylphosphoramidite chemical method on an Applied Biosystems (Foster City, CA) model 381A DNA synthesizer or were purchased from Eurogentech (Seraing, Belgium).

RESULTS

Determination of the Length of Amplifiable Fragments on Cosmid and Genomic DNA

In the framework of the Yeast Genome Sequencing Programme of the European Community, our laboratory was provided with cosmid 14-20. This cosmid has been completely sequenced.[10,11] Because ~40% of the sequence was determined by primer walking, specific primers were available with primer annealing sites scattered over the entire insert. In separate reactions, amplifications were performed both on cosmid DNA and genomic DNA using a common forward primer (646) and a series of ever more distantly located backward primers (584, 574, 561, EG04, EG05, EG08, 583, 578, 566, 680, EG09, 682, 647, 681, EG11, and EG12; see also Fig. 1). On cosmid DNA fragments of the expected sizes (297, 648, 1330, 1578, 2492, 3581, 3775, 4072, 4710, 4851, 5192, 5732, 6432, 7393, 8019, and 8396 bp, respectively) were obtained for all amplification reactions (Fig. 2). With genomic DNA, fragments up to 5.2 kb were obtained. The yield of PCR products, however, was less than with cosmid DNA, especially for larger PCR products (>4 kb). Both the smaller size of amplifiable products and the lower yield on genomic DNA are caused by the higher complexity of genomic DNA versus isolated cosmid DNA. Upon amplification of large fragments (>5 kb), some amplification reactions yielded extra (a-specific) products. A-specific products were probably caused by the presence of a secondary primer annealing site for the backward primer because a secondary annealing site for the forward primer should result in a common a-specific product for all amplification reactions with a backward primer annealing site beyond that secondary annealing site. The latter event was never observed.

Colinearity

Yeast genomic PCR fragments up to 5 kb align well with their corresponding cosmid PCR counterparts (Fig. 2). This result indicates colinearity of the cosmid sequence with the genomic DNA sequence between the annealing positions of primer 646 (at position 610 in the cosmid) and primer EG09 (at position 5802). Twelve primer annealing sites

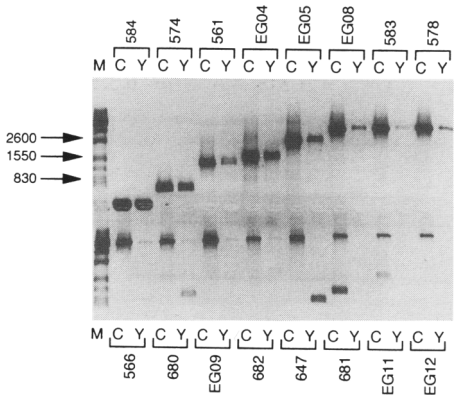


FIGURE 2 PCR products obtained by amplification of both cosmid (C) and yeast genomic (Y) DNA with a common forward primer (646) and ever more distant backward primers. The number of the backward primer is indicated at the *top* (*upper* panel) or at the *bottom* (*lower* panel). The expected sizes of the PCR products are indicated in the text. The sizes of some bands of the molecular weight marker (M) are indicated at *left*.

were used to cover this distance (5192 bp), resulting in an average resolution of 433 bp. For the assessment of colinearity between yeast genomic DNA and the entire cosmid, 18 forward primers were selected separated by ~1800 bp each. Each forward primer was used both on genomic and cosmid DNA in separate amplification reactions with three backward primers, the proximal one being the most distant backward primer of the upstream forward primer (Fig. 3). The result is shown in Figure 4. On both DNA templates, PCR products of the expected sizes were obtained for all amplifications (Table 1). With some amplification reactions small amounts of secondary PCR products were observed. The specific PCR product, however, always produced the major band. Apart from the sequences at the boundaries of the fragments, the entire cosmid DNA (from position 483 to 32,266) has been analyzed and displays colinearity with the yeast

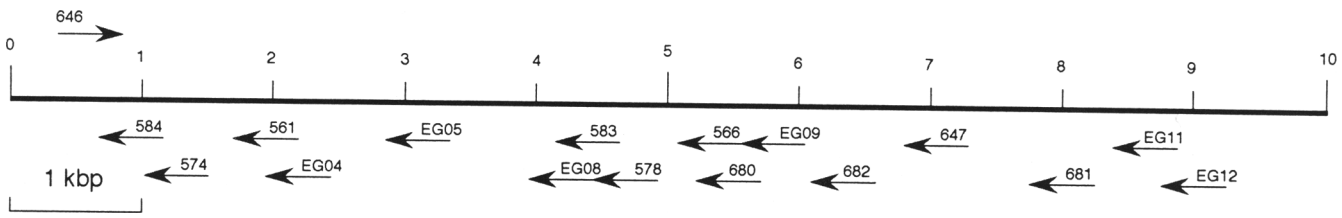


FIGURE 1 Localization and orientation of the primers used for determination of the length of amplifiable products on cosmid and genomic DNA.

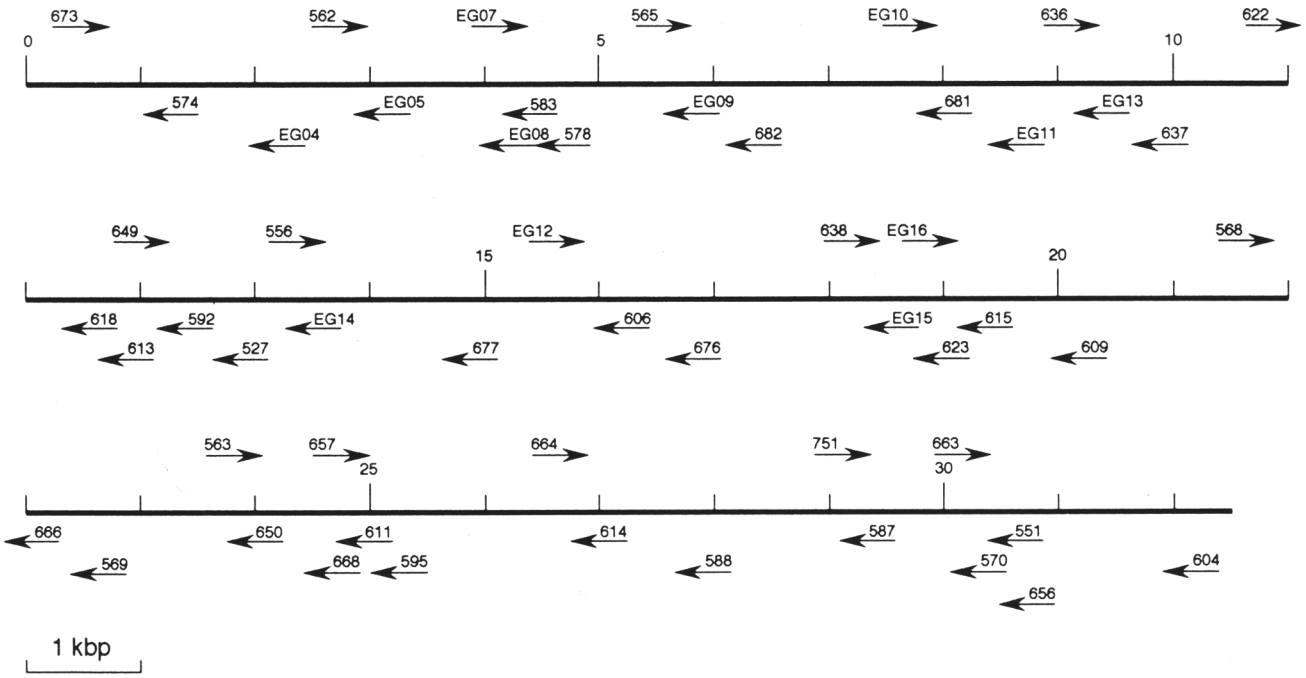


FIGURE 3 Localization and orientation of the primers used for the assessment of colinearity between cosmid and genomic DNA sequences. Forward primers, from *left* to *right*; backward primers, from *right* to *left*.

genomic DNA with an overall resolution of 578 bp (55 primer annealing sites tested over 31,783 bp).

DISCUSSION

Because genome sequencing strategies are usually based on collating sequences from a collection of cloned cosmid or phage DNAs, the integrity of the intermediate large fragments in those vectors must be verified. So far, this is usually done by physical mapping. Comigration on agarose gel electrophoresis of restriction fragments from template and subcloned DNA is analyzed. The resolution of this approach is variable and limited to the distribution of restriction sites and size of the restriction fragments obtained (usually between 1 and 10 kb). Rearrangements within restriction fragments are not detected easily.

In this study we designed a colinearity test based on multiple PCR reactions. We demonstrated colinearity of the cosmid 14-20 DNA sequence with the corresponding sequence on *S. cerevisiae* chromosome XIV by PCR. Both genomic and cosmid DNAs were amplified in 108 separate reactions as outlined above and in Figure 3. All primers were available because the cosmid was sequenced, in part, with walking primers. All PCR products from cosmid DNA comigrated with corresponding PCR products from genomic DNA on agarose gel electrophoresis, indicating colinearity of the sequences from both templates. PCR products ranged in size from 123 to 3146 bp. In this particular experiment, 55 primer annealing sites were tested over a total length of 31,783 bp, resulting in an average resolution of 578 bp. It is very unlikely that rearrangements between primer annealing sites would pass undetected, as the distance between primer annealing sites is small, and only rearrangements that would not change the size of the fragments would remain unnoticed. Resolution is limited by the availability of primers only, so that an even larger set of primers can be designed. Therefore, this approach is an efficient, rapid, and cost-effective tool for verifying colinearity at a desired level of high resolution between a sequenced DNA fragment and the genomic DNA, especially when a substantial part of the sequencing was performed with walking primers.

On the basis of the experimental observation that fragments >5 kb can be amplified on yeast genomic DNA as well as on cosmid DNA, a further reduction in the number of amplification reactions can be conceived. Forward primers could

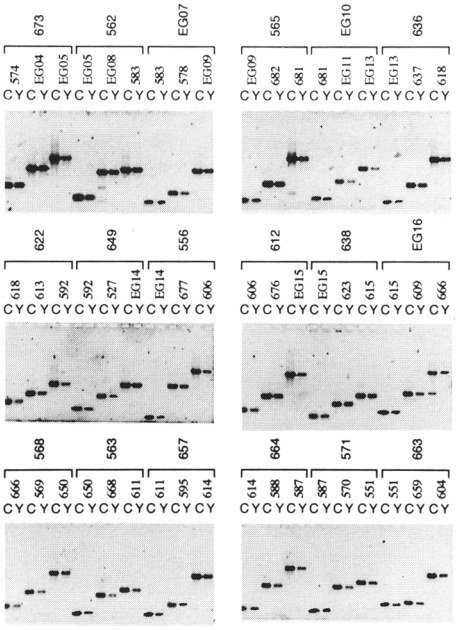


FIGURE 4 PCR products obtained by amplification of both cosmid (C) and yeast genomic (Y) DNA with 18 forward primers, each with three backward primers in separate reactions. Forward primers and backward primers used are indicated at the *top* of each panel. In the reaction with primers EG16 and 666 on cosmid DNA, an a-specific product might have been caused by overflow contamination from the neighboring reaction sample in the gel.

TABLE 1 Expected Sizes of PCR Products Using the Indicated Forward and Backward Primers for the Assessment of Colinearity of Cosmid DNA Sequence with Genomic DNA Sequence

Forward	Backward	Size (bp)	Forward	Backward	Size (bp)
673	574	775	565	EG09	218
	EG04	1705		682	758
	EG05	2619		681	2419
562	EG05	346	EG10	681	283
	EG08	1438		EG11	909
	583	1629		EG13	1649
EG07	583	241	636	EG13	236
	578	538		637	750
	EG09	1658		618	2407
622	618	655	612	606	553
	613	978		676	1174
	592	1500		EG15	2887
649	592	365	638	EG15	321
	527	847		623	765
	EG14	1481		615	1147
556	EG14	123	EG16	615	460
	677	1485		609	1275
	606	2814		666	3146
568	666	401	664	614	314
	569	996		588	1220
	650	2375		587	2653
563	650	160	571	587	209
	668	824		570	1179
	611	1101		551	1490
657	611	178	663	551	439
	595	484		656	541
	614	2229		604	2099

be selected at 2.5 kb from each other. Omitting the internal backward primers (this also reduces the resolution), colinearity analysis of cosmid 14-20 with yeast chromosome XIV would require only twice 26×2 PCR reactions (i.e., 13 forward primers and 14 backward primers). On the same basis, analysis of the entire chromosome XIV (820 kb) is feasible with 656×2 amplification reactions (these can be done in 14 microtiter plates) using 328 forward primers and 329 backward primers. Current developments in long-distance PCR[12] will undoubtedly enable a further reduction in the number of amplification reactions and therefore increase coverage of larger contiguous sequences. Thus, application of this approach to more complex genomes such as the human genome could become feasible.

A useful extension of this method is the rapid and simultaneous analysis of different yeast strains to test diversity among strains in a selected part of the genome. In addition, conservation of functionally interesting sequences among related organisms is readily performed. Sequence diversity would be scored by the absence of PCR product (no primer annealing site) or by the presence of PCR products of different sizes (secondary primer binding site or altered genome structure). Furthermore, both different and identical PCR products from the tested organisms can be sequenced directly.[13,14] Sequencing of PCR products with identical sizes allows for the determination of minor sequence differences, for example, base substitutions.

ACKNOWLEDGMENTS

We thank Dr. P. Philippsen (Basel, Switzerland) for the cosmid 14-20 DNA. The Yeast Genome Sequencing Project is financed by the EC-BIOTECH Programme This work is also supported in part by a grant from the Belgian Federal Services of Science Policy. P.V. is a recipient of a postdoctoral fellowship of the Catholic University Leuven, Belgium.

REFERENCES

1. Hood, L., B. Koop, J. Goverman, and T. Hunkapiller. 1992. Model genomes: The benefits of analysing homologous human and mouse sequences. *Trends Biotechnol.* **10:** 19–22.
2. Magnien, E., M. Bevan, and K. Planqué. 1992. A European "BRIDGE" project to tackle a model plant genome. *Trends Biotechnol.* **10:** 12–15.
3. Smith, L. and L. Hood. 1987. Mapping and sequencing the human genome: How to proceed. *Biotechnology* **5:** 933–939.
4. Vassarotti, A. and A. Goffeau. 1992. Sequencing the yeast genome: The European effort. *Trends Biotechnol.* **10:** 15–18.
5. Wilson, R. 1994. 2.2 Mb of contiguous nucleotide sequence of chromosome III of *C. elegans*. *Nature* **368:** 32–38.
6. Sternberg, N. 1990. Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proc. Natl. Acad. Sci.* **87:** 103–107.
7. Burke, D.T., G.F. Carle, and M.V. Olson. 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236:** 806–812.
8. Evans, G.A. and G.M. Wahl. 1987. Cosmid vectors for genomic walking and rapid restriction mapping. *Methods Enzymol.* **152:** 604–610.
9. Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:** 487–491.
10. Verhasselt, P., R. Aert, M. Voet, and G. Volckaert. 1994. Nucleotide sequence analysis of an 8887 bp region of the left arm of yeast chromosome XIV, encompassing the centromere sequence. *Yeast* **10:** (in press).
11. Verhasselt, P., R. Aert, M. Voet, and G. Volckaert. 1994. Twelve open reading frames revealed in the 23.6 kb segment flanking the centromere on the *Saccharomyces cerevisiae* chromosome XIV right arm. *Yeast* **10:** 945–951.
12. Cohen, J. 1994. "Long PCR" leaps into larger DNA sequences. *Science* **263:** 1564–1565.
13. Verhasselt, P., S. Van Campenhout, M. Voet, B. Wellens, and G. Volckaert. 1993. A general method for routine sequencing of cloned DNA fragments using commercial dye primers and its application in sequence analysis of *Toxoplasma gondii* RH genome fragments. *J. DNA Sequencing Mapping* **4:** 71–77.
14. Hultman, T., S. Stahl, E. Hornes, and M. Uhlén. 1989. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads. *Nucleic Acids Res.* **17:** 4937–4946.

Received April 22, 1994, accepted in revised form July 6, 1994.