Prediction and identification of recurrent genomic rearrangements that generate chimeric chromosomes in *Saccharomyces cerevisiae*

Kim Palacios-Flores[a], Alejandra Castillo[a], Carina Uribe[a], Jair García Sotelo[a], Margareta Boege[a], Guillermo Dávila[a], Margarita Flores[a], Rafael Palacios[a,1], and Lucia Morales[a,1]

[a]Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Juriquilla, Querétaro 76230, México

Contributed by Rafael Palacios, March 11, 2019 (sent for review November 16, 2018; reviewed by Charles Lee and Xavier Perret)

Genomes are dynamic structures. Different mechanisms participate in the generation of genomic rearrangements. One of them is nonallelic homologous recombination (NAHR). This rearrangement is generated by recombination between pairs of repeated sequences with high identity. We analyzed rearrangements mediated by repeated sequences located in different chromosomes. Such rearrangements generate chimeric chromosomes. Potential rearrangements were predicted by localizing interchromosomal identical repeated sequences along the nuclear genome of the *Saccharomyces cerevisiae* S288C strain. Rearrangements were identified by a PCR-based experimental strategy. PCR primers are located in the unique regions bordering each repeated region of interest. When the PCR is performed using forward primers from one chromosome and reverse primers from another chromosome, the break point of the chimeric chromosome structure is revealed. In all cases analyzed, the corresponding chimeric structures were found. Furthermore, the nucleotide sequence of chimeric structures was obtained, and the origin of the unique regions bordering the repeated sequence was located in the expected chromosomes, using the perfect-match genomic landscape strategy (PMGL). Several chimeric structures were searched in colonies derived from single cells. All of the structures were found in DNA isolated from each of the colonies. Our findings indicate that interchromosomal rearrangements that generate chimeric chromosomes are recurrent and occur, at a relatively high frequency, in cell populations of *S. cerevisiae*.

genomic rearrangements | chimeric chromosomes | reciprocal translocations | genome architecture | PMGL strategy

Since the pioneering studies of Barbara McClintock (1), genomes have been established as dynamic structures. Several mechanisms participate in the generation of genomic rearrangements, all of them based on central biological processes such as transposition (2); DNA repair, including NAHR and nonhomologous end joining (3, 4), and DNA replication, including break-induced replication, fork stalling, and template switching; and origin-dependent inverted-repeat amplification (5, 6). Genomic rearrangements have been associated with pathological conditions, including genomic disorders (7) and cancer (8); with resistance to drugs (9); and with adaptive processes (10). The field of experimental evolution (11) has also provided examples of the central biological role of genomic rearrangements (12). Furthermore, genomic rearrangements are the source of structural variation. Although structural variants were initially considered only in the context of pathological conditions, pioneering experiments from two independent groups indicated that structural variation is present in the genome of normal humans (13, 14).

The present study is focused on genomic rearrangements generated by NAHR. The targets of recombination in NAHR are long repeated sequences with high identity. According to the position and orientation of the repeated sequences, different types of rearrangements can be generated by this mechanism. Repeated sequences present in the same DNA molecule in direct orientation can generate duplications (which in turn can lead to amplifications) or deletions of genetic material. If the repeated sequences are located in inverted orientation, inversions of genetic material can be generated. Repeated sequences located in different circular molecules, such as bacterial circular chromosomes or plasmids, can generate cointegration of the corresponding replicons. If the repeated sequences are located in different linear DNA molecules, such as eukaryotic chromosomes, NAHR might generate chimeric molecules, harboring the proximal region of one chromosome and the distal region of another chromosome. Most interesting, if the position and orientation of the repeated sequences present along a genome are known, it is possible to predict the different types of potential rearrangements that can be generated by NAHR.

Using a PCR-based strategy, we previously detected the products resulting from different types of rearrangements generated by NAHR. In the *Rhizobium* genome we analyzed duplication-amplifications and deletions (15) as well as replicon cointegrations (16); in the human genome, we analyzed inversions of genetic material (17). In the present study, we predicted, identified, and quantified interchromosomal rearrangements that generate chimeric chromosomes in the yeast *Saccharomyces cerevisiae*. This organism represents the most used model of a simple eukaryote. It has several advantages, including a compact genome of about 12 Mb, that has been accurately sequenced,

Significance

Genome dynamics has implications in different biological phenomena, including pathological and evolutionary processes. Nonallelic homologous recombination generates genomic rearrangements mediated by long repeated sequences with high identity. By locating repeated regions shared by two different chromosomes along the genome of the yeast *Saccharomyces cerevisiae*, we predicted rearrangements that generate chimeric chromosomes. A PCR-based strategy revealed the chimeric structures predicted. The alignment of the nucleotide sequence of the PCR products to the genome ascertained the expected origin of the chimeric structures. Such structures were found in the culture and in colonies isolated from single cells. Our findings indicate that interchromosomal rearrangements that generate chimeric chromosomes are recurrent and occur, at a relatively high frequency, in cell populations of *S. cerevisiae*.

Author contributions: K.P.-F., M.B., G.D., M.F., R.P., and L.M. designed research; K.P.-F., A.C., C.U., M.F., and L.M. performed research; K.P.-F. and J.G.S. contributed new reagents/analytic tools; J.G.S. and R.P. analyzed data; and R.P. and L.M. wrote the paper.

Reviewers: C.L., The Jackson Laboratory for Genomic Medicine; and X.P., University of Geneva.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: palacios@liigh.unam.mx or lmorales@liigh.unam.mx.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1819585116/-/DCSupplemental.

Published online April 8, 2019.

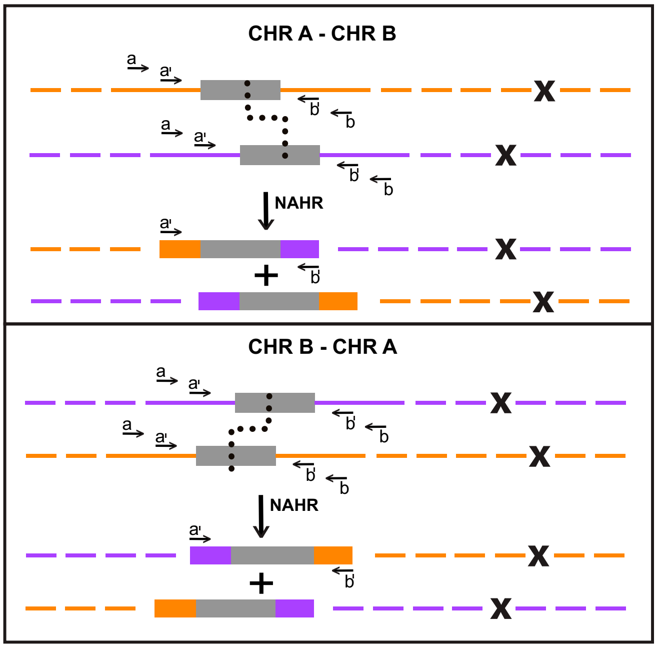

Fig. 1. Schematic of the generation and identification of chimeric chromosomes. Two chromosomes sharing an identical repeated sequence are shown. Orange, chromosome A (CHR A); purple, chromosome B (CHR B); gray, repeated sequence; X, centromere. The solid line indicates a region of unique sequences close but outside the repeated region, where the PCR primers must be located; the broken line indicates a fragment of the rest of the chromosomes. Short arrows indicate the relative position of the PCR primers used to detect the wild-type and chimeric structures. Forward primers, a and a^{I}; reverse primers, b and b^{I}. The dotted black line indicates the site where NAHR generated two chimeric chromosomes. If forward primers located in one chromosome and reverse primers located in the other chromosome are used, the chimeric structures are detected. The corresponding structures can be detected either from CHR A–CHR B (*Top*) or from CHR B–CHR A (*Bottom*) according to the positions of the forward and reverse primers.

annotated, and systematically mutated; a fast division rate; the possibility of propagation in either haploid or diploid state; and highly sophisticated tools for genetic manipulation. We conclude that interchromosomal rearrangements driven by NAHR, that generate chimeric chromosomes, are recurrent and occur, at a relative high frequency, in the model yeast organism *Saccharomyces cerevisiae*.

Results

Rationale for the Prediction and Identification of Interchromosomal Genomic Rearrangements. High-identity repeated sequences located in different chromosomes can be targets of NAHR. Such recombination generates interchromosomal genomic rearrangements which in turn produce chimeric chromosomes. Fig. 1 schematizes this phenomenon. Two chromosomes, A and B, share a repeated sequence. NAHR between them results in the simultaneous generation of two chimeric chromosomes, with reciprocal translocations. These rearrangements can be predicted in the whole genome by locating the corresponding repeated regions shared by pairs of different chromosomes. To detect the generation of chimeric chromosomes, we used a PCR-based strategy. PCR primers corresponding to a unique region located near the repeated region of each chromosome were designed and synthesized. When PCR is performed using a forward primer from chromosome A and a reverse primer from chromosome B (CHR A–CHR B) or a forward primer from chromosome B and a reverse primer from chromosome A (CHR B–CHR A), the resulting PCR products contain the break points of the corresponding chimeric chromosome.

Location of Interchromosomal Repeated Sequences in the *S. cerevisiae* Genome. The nuclear reference genome of *S. cerevisiae* strain S288C was analyzed to locate high-identity repeated sequences shared by pairs of different chromosomes. According to the reported nucleotide sequence of the strain (18, 19), we searched for identical sequences equal to or larger than 1 kb located in different chromosomes (see *Materials and Methods*). NAHR requires a minimum of perfect homology, on the order of 300 nucleotides. To have optimal targets for NAHR, we selected regions with large tracts of identity—in this case, at least 1 kb. A total of 164 such pairs of sequences were found. For each pair of sequences, *SI Appendix*, Table S1,` reports the chromosomes involved, the position in each chromosome, the size of the identical sequence, and a brief summary of the annotation. Fig. 2 shows the 16 chromosomes of *S. cerevisiae* arranged as a circle, with colored lines connecting the corresponding pairs of repeated elements considered in this work. We distinguished two types of pairs: those where both repeats are located either in the left arm or in the right arm of the respective chromosomes (133 pairs; orange or green lines); and those where one repeat is located in the right arm and the other in the left arm of the two chromosomes (31 pairs; purple lines). Following a NAHR event, the former pairs might generate two viable chimeric chromosomes (Fig. 1). In contrast, the latter pairs would generate one chromosome with two centromeres and one chromosome without centromeres. Such structures should be nonviable unless secondary rearrangements occur (20), and those were not considered in the present study. We focused on analysis of the results of NAHR in the pairs of identical sequences shown as green lines in Fig. 2 and highlighted in the *SI Appendix*, Table S1. Six pairs of sequences were randomly selected, and all of them involved transposons.

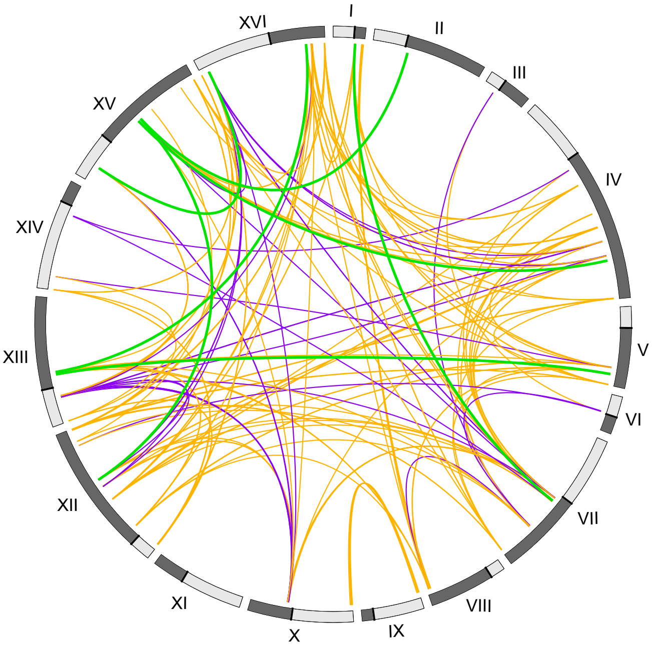

Fig. 2. Location of interchromosomal repeated sequences. Identical repeated sequences larger than 1 kb and shared by different chromosomes were located (*Materials and Methods*). The chromosomes were arranged as a circle (I–XVI). The position of the centromere is indicated by a black line; the left arm, light gray line; the right arm, dark gray line. Lines joining the corresponding chromosomes indicate the position of each pair of repeated sequences. Orange and green lines join repeated sequences located both either in the left arm or in the right arm of the corresponding chromosomes; green lines indicate the subset analyzed. Purple lines join repeated sequences, one located in the left arm and the other in the right arm of the corresponding chromosomes.

One pair of sequences, corresponding to elongation factor 2 paralogous genes (EFT1/EFT2), was intentionally added (see *SI Appendix*, Table S1).

Identification of Chimeric Chromosome Break Points. As schematized in Fig. 1, the presence of chimeric chromosomes in an *S. cerevisiae* culture was detected by identifying the corresponding break points generated by NAHR between pairs of chromosomes. Forward and reverse PCR primers were generated from the unique sequence adjacent to the repeated sequence in each chromosome. It should be pointed out that the identical repeated sequences reported in the *SI Appendix*, Table S1 are usually framed by additional repeats both upstream and downstream, located in the corresponding chromosome and/or elsewhere in the genome. In particular, most of such sequences correspond to transposons (see *SI Appendix*, Table S1) that have a size of about 6 kb and that are highly repeated throughout the genome. The PCR primers should border the complete repeated region and must be located in a unique region, taking into account the whole genome. In the present study, such regions were localized using the perfect-match genomic landscape (PMGL) strategy (21) (see *Discussion* and *Material and Methods*). If the forward and reverse primers are located in the same chromosome, a wild-type (nonrearranged) structure will be detected. In contrast, if the forward primer is located in one chromosome and the reverse primer is located in another chromosome, a chimeric structure harboring the break point of the NAHR will be detected. To increase PCR specificity, we performed nested PCRs (see *Materials and Methods*).

PCR products from wild-type and chimeric structures are shown in the *SI Appendix*, Fig. S1 *A* and *B*, respectively. The reactions involving transposons usually generate two fragments, one large and one small. The large fragment harbors the complete wild-type or chimeric transposon; the small fragment results from a deletion within the transposon (see *Discussion*). As expected, the PCRs that do not involve transposons, either wild-type (*SI Appendix*, Fig. S1*A*, lanes 9 and 10) or chimeric (*SI Appendix*, Fig. S1*B*, lanes 5 and 6) yield only one PCR product. Data relevant to wild-type structures and chimeric structures are shown in the *SI Appendix*, Tables S2 and S3, respectively.

Identification of the Genomic Origin of PCR Products Harboring Chimeric Chromosome Break Points. The generation of PCR products using forward primers from one chromosome and reverse primers from another chromosome strongly suggests the presence of chimeric chromosome structures in the cell populations analyzed. To corroborate the chimeric nature of such structures, the nucleotide sequence of several PCR products was obtained (see *Materials and Methods*) and compared with the nucleotide sequence of the corresponding region(s) of the chromosome(s) involved in the generation of the PCRs. To this end, a methodology based on the PMGL strategy (21) was used (see *Materials and Methods*). As shown in Fig. 3, the PMGLs are generated by merging a reference genome string dataset, from the region of the reference genome corresponding to the PCR, with a sequence reads dataset from the PCR product. The PMGLs reflect the coverage of sequence reads along the different regions of the PCR product and reveal their corresponding origins. All of the larger fragments of the PCR products shown in the *SI Appendix*, Fig. S1, as well as some of the small fragments, were sequenced and PMGLs were generated. The PCR identifiers of the products generated in this study and the identifiers of the PMGLs of all PCR products sequenced are presented in the *SI Appendix*, Tables S2 and S3, respectively.

Fig. 4 shows an example of the analysis of PCR products generated from a wild-type region located in chromosome I. The PCR generated two products, a large fragment of 6,334 nucleotides and a small fragment of 746 nucleotides (see *SI Appendix*, Table S2). The large fragment of the PCR (Fig. 4*A*) contains the sequence of the transposon, as well as the unique regions of chromosome I that border the transposon. A PMGL was generated using the reference string dataset from the corresponding region of chromosome I and the sequence read string dataset from the corresponding PCR product. The PCR sequence reads cover all regions of the PCR. As expected, the regions of the reference genome that are outside the region corresponding to the PCR product did not attract read strings. The small PCR fragment was also sequenced (Fig. 4*B*), and the PMGL was generated from the same region of chromosome I used for the large fragment and the set of read strings from the small fragment. The proximal and distal regions of the PMGL show the same characteristics as the PMGL derived from the large fragment. In contrast, a large region of about 6 kb in the zone of the transposon does not attract any read strings (*Discussion*).

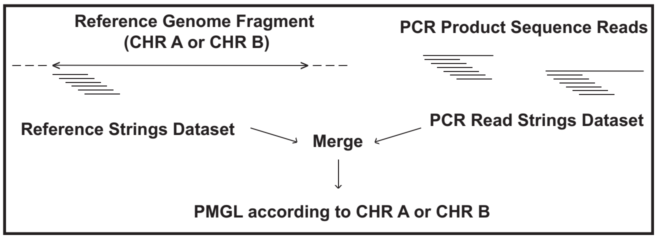
Reference Genome Fragment
(CHR A or CHR B)
PCR Product Sequence Reads
Reference Strings Dataset
PCR Read Strings Dataset
Merge
PMGL according to CHR A or CHR B

Fig. 3. Genomic location of the PCR fragments derived from wild-type and chimeric structures. Illumina reads from each PCR fragment were obtained (see *Materials and Methods*). A fragment of the reference genome was cut in the corresponding chromosome (wild-type structure) or chromosomes (chimeric structures) PMGL was used to determine the coverage of the PCR sequence reads along the corresponding reference genome fragment (see *Material and Methods*). The reference genome fragment is shown as a line bordered by arrows. This fragment is divided into overlapping 25-nucleotide strings shifted by 1 nucleotide. This constitutes the ordered reference string dataset. The rest of the chromosome is shown as a broken line. Each sequence read from the PCR product is similarly divided. This constitutes the PCR read string dataset. Both datasets are merged, and a directed PMGL is generated to indicate the coverage of PCR read strings at each ordered position of the reference genome.

Fig. 5 shows an example of the PCR products generated by NAHR between repeated sequences shared by chromosomes II and XV. In this case, the forward primers were located in chromosome II and the reverse primers in chromosome XV (II–XV). Since the NAHR involves transposons in the two chromosomes, two PCR fragments were produced: a large fragment (Fig. 5*A*) and a small fragment (Fig. 5*B*). The II–XV PCR products contain, from the proximal to the distal end, a unique region of chromosome II, the transposon sequences (in the case of the large fragment) or a fragment of the transposon sequences (in the case of the small fragment), and a unique region of chromosome XV. When the PMGL was made using the sequence reads from the respective fragment and the region of the reference genome from chromosome II, read strings were attracted from the regions of the PCR corresponding to the unique region of chromosome II, the transposon sequences (large fragment) or a fragment of the transposon sequences (small fragment), but not from the unique region corresponding to chromosome XV. In contrast, when the PMGL was generated using as a reference a fragment from chromosome XV, the unique region of chromosome II did not attract read strings. The reaction was also performed locating the forward primers in chromosome XV and the reverse primers in chromosome II (XV–II; see *SI Appendix*, Fig. S2). The PCR products presented a general pattern similar to that in Fig. 5.

The PCR products corresponding to the chimeric structures generated by NAHR between repeated regions harboring elongation factor 2 paralogous genes located in chromosomes IV and XV are shown in the *SI Appendix*, Fig. S3. The reaction was primed either IV–XV (*SI Appendix*, Fig. S3A) or XV–IV (*SI Appendix*, Fig. S3B). When the corresponding PMGLs were generated using a reference region from chromosome IV, the unique region corresponding to chromosome XV did not attract read strings. In contrast, when the PMGLs were generated

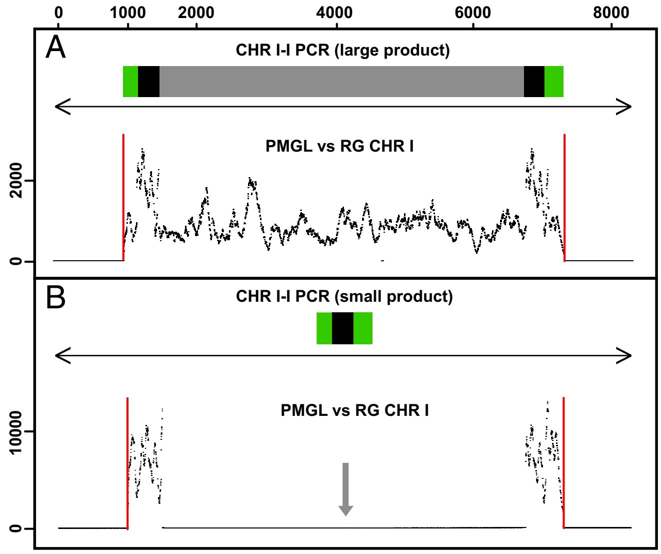

Fig. 4. Origin of the PCR products generated by a wild-type structure involving a transposon. The data were derived from a PCR generated by forward and reverse primers located near but outside the repeated region in chromosome I (CHR I-I). The PCR (ID 2702; *SI Appendix*, Table S2) shows two products: (*A*) a large product and (*B*) a small product (*SI Appendix*, Fig. S1*A*). The products are schematized as rectangles. Green, regions upstream and downstream of the transposon (see *Results*) that correspond to unique sequences in the chromosome; black, regions corresponding to the transposon direct repeated regions or delta sequences; gray, the rest of the transposon or epsilon sequence; line bordered by two-headed arrows, fragment cut from the reference genome; red bars, limits of the large PCR product. The PMGL was generated from the ordered reference string dataset (scale in nucleotides at the top) and the PCR read string dataset from either the large product (PMGL 01_01) or the small product (ID PMGL 027_01) (see *SI Appendix*, Table S2); coverage of read strings is presented on the *y*-axis of the corresponding panel. The gray arrow indicates the zone of the transposon that did not attract read strings.

using a reference genome from chromosome XV, the unique region corresponding to chromosome II did not attract reference strings.

Estimation of the Relative Concentration of Chimeric Structures Versus Wild-Type Structures. The data just presented indicate that the corresponding PCR products represent either wild-type or chimeric chromosome structures. The concentration of chimeric structures relative to wild-type structures was estimated (see *Discussion*). To this end, serial dilutions of the DNA from a particular cell population were made and the corresponding PCRs were performed for each dilution. The comparison of the highest dilution at which wild-type structures were observed with that at which chimeric structures were observed was an approximation of the relative amount of chimeric vs. wild-type structures (see *Materials and Methods*).

A DNA sample from a culture of *S. cerevisiae* strain S288C, or from colonies derived from single cells (see below) corresponding to about 105 cells, was serially diluted by a factor of 2. PCRs primed to detect either wild-type structures or chimeric structures were performed in each dilution. PCRs were performed to detect wild-type structures from chromosomes I, II, IV, V, VII, XII, XV, and XVI; chimeric structures, from chromosomes I–VII, XV–II, XV–IV, XIII–V, XV–XII, XIII–XVI, and XVI–XV. Fig. 6 shows an example of the PCR products obtained from a wild-type structure from chromosome VII (*A*) and a chimeric structure from chromosomes I–VII (*B*). Much higher dilutions show the wild-type structure compared with those showing the chimeric structure. Interestingly, the different wild-type structures analyzed generated products up to about the same dilution (not shown), while the chimeric structures varied among them (see *Discussion* and Fig. 7).

Generation of Chimeric Chromosomes in Colonies Derived from Single Cells. A highly diluted culture of S288C *S. cerevisiae* was plated to obtain isolated cells that formed independent colonies. DNA was isolated from each of four such colonies, and PCRs were performed using primers to detect the following chimeric chromosomes: I–VII, XV–II, XV–IV, XIII–V, XV–XII, XIII–XVI, and XVI–XV. As shown in Fig. 7, each of the seven chimeric chromosome structures was present in the culture and in each of the four colonies. The relative concentration of chimeric vs. wild-type structures was estimated in the culture and in the different colonies as explained previously (Fig. 6). The relative concentrations of chimeric structures varied according to both the origin of the DNA and the specific rearrangement (see *Discussion*).

Discussion

By analyzing the nucleotide sequence of the reference genome of *S. cerevisiae* strain S288C and locating identical sequences shared by different chromosomes, we could predict potential rearrangements

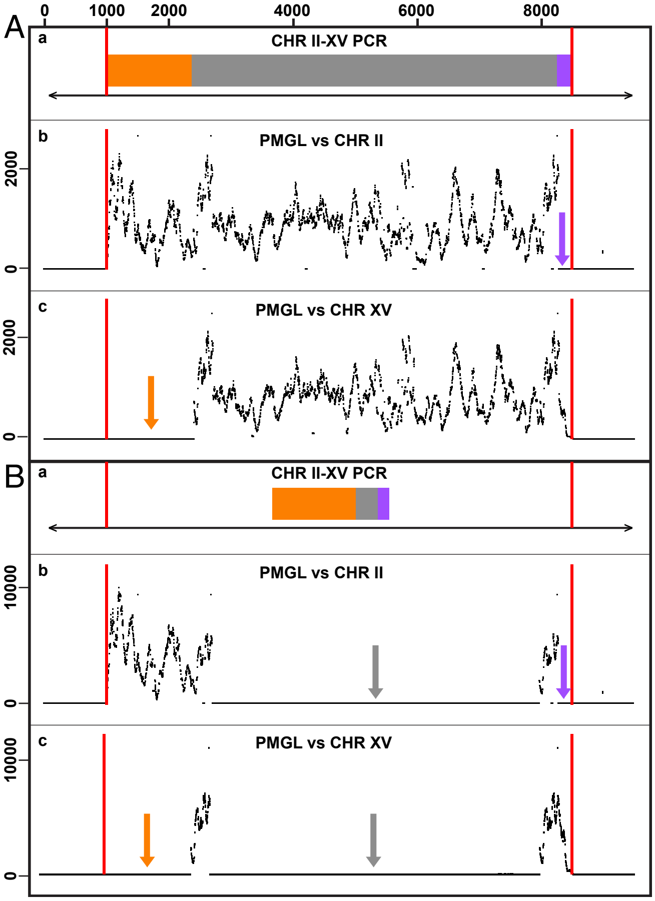

Fig. 5. Origin of the PCR products generated by a chimeric structure involving transposon sequences. The PCR was primed with forward primers located in chromosome II and reverse primers located in chromosome XV. The PCR (ID 2401; see *SI Appendix*, Table S3) produced a large fragment (*A*) and a small fragment (*B*). The products are schematized as rectangles (*A, a*; *B, a*) and contain a unique region of chromosome II (orange): the transposon sequences (large fragment) or a fragment of the transposon sequences (small fragment; gray); and a unique region of chromosome XV (purple). Shown in *A*, *b*; *A, c*; *B*, *b*; and *B*, *c* are the PMGLs obtained using as reference a fragment of the indicated chromosome. Orange arrows, regions that did not attract read strings from chromosome II; purple arrows, regions that did not attract read strings from chromosome XV; gray arrows, regions of the transposon sequences that did not attract read strings from the small fragment. Other indications are as in Fig. 4.

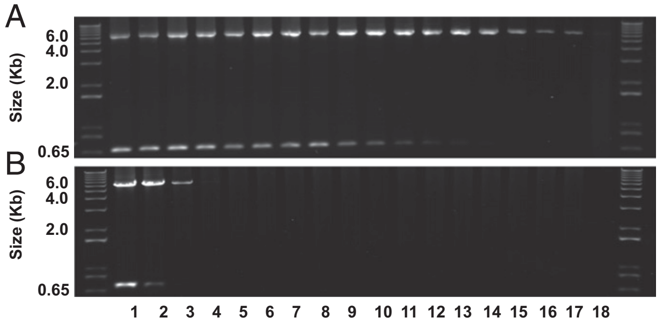

Fig. 6. Estimation of the proportion of chimeric structures relative to wild-type structures. DNA was serially diluted by a factor of 2. Each dilution was used to generate a PCR from either a wild-type structure (here CHR VII, ID PCR 2710; see *SI Appendix*, Table S2) or a chimeric structure (here CHR I-CHR VII, ID PCR 2382; see *SI Appendix*, Table S3). Agarose gel electrophoresis was used to detect the corresponding PCR products. (*A*) wild-type structure; (*B*) chimeric structure. The first and last lanes of the gels show a reference ladder, with the corresponding sizes in kb shown in the *y*-axis of each gel. The comparison of the maximal dilution at which chimeric structures are observed with the maximal dilution at which wild-type structures are observed is an indication of the relative number of chimeric structures present in the population. In this case, there are about 2^3 chimeric structures per ~2^{17} wild-type structures.

produced by the NAHR mechanism. Such rearrangements generate chimeric chromosomes. Most of the repeated sequences found as potential targets for interchromosomal rearrangements corresponded to transposon sequences. The structure of each yeast transposon consisted of two direct repeated regions of about 300 bp (delta sequences), which bordered a unique sequence of about 6 kb (epsilon sequence) (22). Whole Ty elements were present as interspersed repeated sequences along the 16 chromosomes of the *S. cerevisiae* genome. The dynamics of NAHR involving transposon sequences is complex. The two delta sequences of a Ty element can internally recombine, producing either a deletion of the epsilon sequence of the transposon or a tandem duplication-amplification of the whole transposon sequence. The deletion of the epsilon sequence could explain the small fragment observed in the PCR products of either wild-type or chimeric structures observed throughout the study (Figs. 4 and 5 and *SI Appendix*, Figs. S1 and S2). Repeated regions of the transposon or solo delta sequences present in different regions of the genome can be targets of NAHR recombination as well. If they are present in the same chromosome, deletions, duplications-amplifications, or inversions can be produced. If they are located in different chromosomes (as analyzed in this work) they can generate chimeric chromosomes. Furthermore, chimeric chromosomes can in turn recombine with other transposons or solo delta sequences located in different places along the genome.

The break points expected to be present in the chimeric chromosomes were analyzed using a PCR-based approach. Of the seven rearrangements analyzed, six mediated by transposon sequences and one by elongation factor 2 paralogous genes (EFT1/EFT2), all of the expected break points were detected. The genomic origin of the chimeric structures present in the PCR products was ascertained by generating the nucleotide sequence of several PCR products and aligning it to the corresponding genomic regions of the chromosomes involved in the rearrangements. This was performed by applying a methodology based on the PMGL strategy. This strategy was recently developed in our laboratory (21) to compare genomes and detect variation using only perfect matches between a reference and a query genome.

The precise quantification of genomic rearrangements generating chimeric chromosomes represents an elusive problem. As explained above, the dynamics of NAHR among transposon sequences is complex. In addition, different primers, which are essential to detect wild-type vs. chimeric structures, could produce PCR amplifications of different magnitude. As indicated in *Results*, we estimated the relative proportion of wild-type and chimeric structures by performing PCR reactions in serial dilutions of the template DNA, both from the culture and from individual colonies derived from single cells. We found that the wild-type reactions generated products up to about the same dilution, while specific rearrangements varied among them and also according to the source of the template. Our general estimate is that chimeric structures were present, at an average concentration of 10^{-3} in the culture and 10^{-4} in the individual colonies. It can be assumed that the original cell did not contain the rearrangements. Thus, each specific rearrangement should have appeared at some point(s) during the reproduction of the cell. Our data indicate that NAHR-mediated rearrangements that generate chimeric chromosomes are recurrent and occur at relatively high frequency in the genome of the model strain studied.

It is interesting that the generation of genomic rearrangements has been established in yeast in multiple in vivo studies (23–26), including experimental evolution cases where chimeric chromosomes have been generated. Natural evolution also shows that rearrangements have shaped the chromosomal landscape in yeast. The whole genome sequence of more than 1,000 natural isolates of *S. cerevisiae* has been determined (27, 28). The nucleotide sequences of seven strains of *S. cerevisiae* and five strains of its most related species, *Saccharomyces paradoxus*, were generated by assemblies using long-read technologies. It was found that both species show a high level of interchromosomal reshuffling (27). Furthermore, the chromosomal landscape of *S. cerevisiae* has been altered using genome editing (29, 30).

The human genome contains a very large number of interspersed repeated sequences, with the *Alu* sequences the most abundant of all. NAHR between *Alu* sequences has reshaped the primate genome (31) and has generated rearrangements related to genetic diseases and to cancer genomic instability (32). The methodology outlined here provides a simple bioinformatic and experimental strategy to rapidly and accurately predict and identify genomic rearrangements at break-point resolution. Such a strategy could be useful in a variety of research and applied areas, including genome dynamics, experimental evolution, genetic diseases, and cancer.

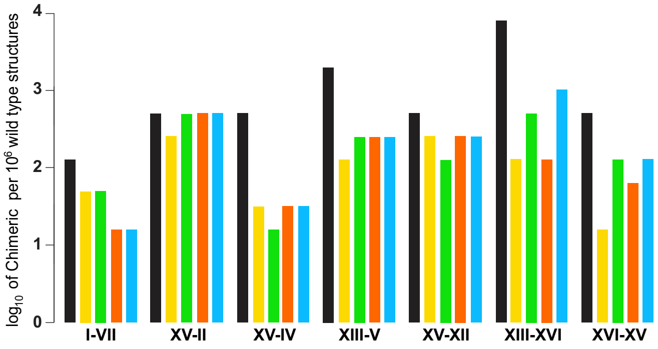

Fig. 7. Identification and estimation of the relative proportion of interchromosomal rearrangements from colonies derived from single cells. The culture of strain *S. cerevisiae* S288C was plated at high dilution to obtain separated cells. Cells were grown to generate single colonies. DNA was extracted from the culture and from each of four colonies. The proportion of different chimeric structures relative to wild-type structures was estimated as in Fig. 6. Bars indicate the relative number of chimeric structures (*y*-axis). Black, culture; yellow, colony 1; green, colony2; orange, colony 3; blue, colony 4. Each chimeric structure is indicated at the bottom: the first chromosome is that containing the forward primers; the second chromosome is that containing the reverse primers.

Materials and Methods

***S. cerevisiae* Strain and Growth Conditions.** Strain S288C was obtained from the American Type Culture Collection. The strain was stored at −80 °C. It was cultured at 30 °C in rich yeast extract–peptone–dextrose (YPD) medium, either in liquid with agitation (250 rpm) or in agar.

DNA Isolation, PCR Conditions, and Illumina Sequencing of PCR Products. DNA was isolated from liquid cultures or colonies using the Yeast DNA Extraction Kit from Thermo Scientific. PCR was performed using the Verity Thermal Cycler from Applied Biosystems. Nested PCRs were performed using primers located in the positions indicated in the *SI Appendix*, Tables S2 and S3; the sequence of each primer is presented in the *SI Appendix*, Table S4. In the first PCR, forward primers, a, and reverse primers, b, were used; in the second PCR, forward primers, a^I, and reverse primers, b^I, were used (Fig. 1). The total volume of both, the first and the second PCR reactions, was 25 μL. The reaction contained the template DNA (see below); each primer at a final concentration of 0.4 μM; and Platinum PCR SuperMix High Fidelity from Invitrogen, which supplied the following reagents: dNTPs at a final concentration of 200 μM; 0.5 U recombinant Taq DNA polymerase (*Pyrococcus* species GB-D thermostable polymerase and Platinum TaqAntibody); 60 mM Tris-SO_4 (pH 8.9); 19 mM $(NH_4)_2SO_4$; 2.0 mM $MgSO_4$; and stabilizers. An additional 1 U of the same polymerase was added to each reaction. The template DNA concentration for the first PCR was 1 ng. After the first reaction, the resulting product was purified using the QIAquick PCR Purification Kit from QUIAGEN and recovered in 25 μL of water. For the second (nested) PCR 1 μL of the recovered product was used. The conditions for both reactions were 94 °C for 1 min, 32 cycles (94 °C for 30 s, 58 °C for 30 s, 68 °C for 6 min), and 72 °C for 10 min. To sequence the PCR products, libraries were prepared using the Nextera XT DNA Library Preparation Kit from Illumina. Sequencing was performed in a NextSeq 500 Mid Output v2 Kit (300 cycles) from Illumina.

Location of Interchromosomal Repeated Sequences. The reference genome of the model strain S288C (18, 19) version R64 was downloaded from the National Center for Biotechnology Information. Using Mummer (33), we located pairs of identical repeated sequences larger than 1 Kb and present in different chromosomes along the whole nuclear genome of *S. cerevisiae*.

Location of Unique Sequences Adjacent to Repeated Sequences. PCR primers must be located in unique regions adjacent to the repeated sequences targeted for searching either wild-type or chimeric regions. Such regions must be unique in regard to the whole genome. To localize these regions, the first module of the PMGL algorithm (20) was used. The reference genome sequence of strain S288C was utilized in FASTA format to generate a binary database with Bowtie (34) and to generate the reference string dataset (25 nucleotides each with a shift of 1 nucleotide). The number of identical strings of each reference string in the whole genome was counted and a reference genome self-landscape (RGSL) was generated (21). The extended region(s) corresponding to the repeated sequence(s) in the corresponding chromosome(s) were analyzed in the RGSL to locate the repeated and unique sequences of interest.

Generation of PMGLs to Identify the Genomic Origin of PCR Products. The sequence of the reference genome corresponding to the PCR product (large fragment when transposons were involved) extended upstream and downstream was cut from the corresponding chromosomes. Each sequence was divided in overlapping strings (25 nucleotides each with a shift of 1 nucleotide) to generate an ordered reference string dataset. Similarly, each sequence read from the corresponding PCR product was divided into strings to generate the read strings dataset. Directed PMGLs (21) were then generated for each PCR product by reporting the coverage of read strings for each reference string (Fig. 3). In the case of wild-type structures, one PMGL between the PCR product sequence and the reference sequence of the corresponding chromosome was generated. In the case of chimeric structures, two PMGLs were generated for each PCR product sequence, using the reference strings dataset from each of the chromosomes involved.

ACKNOWLEDGMENTS. We thank Delfino García for technical assistance and Eglee Lomelin for assistance with the artwork. This work received support from Luis Aguilar, Alejandro de León, and Carlos S. Flores of the Laboratorio Nacional de Visualización Científica Avanzada for providing supercomputer services, and from the Instituto Nacional de Medicina Genómica (INMEGEN) for performing the Illumina sequence of PCR products. This work was supported in part by Grant IA207817 from Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) and Grant 2016-01-284992 from Consejo Nacional de Ciencia y Tecnología (CONACYT).

1. McClintock B (1951) Mutable loci in maize. *Year B Carnegie Inst Wash* 50:174–181.
2. Prak ETL, Kazazian HH, Jr (2000) Mobile elements and the human genome. *Nat Rev Genet* 1:134–144.
3. Stankiewicz P, Lupski JR (2002) Genome architecture, rearrangements and genomic disorders. *Trends Genet* 18:74–82.
4. Haber JE (2000) Partners and pathways repairing a double-strand break. *Trends Genet* 16:259–264.
5. Lee JA, Carvalho CMB, Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 131:1235–1247.
6. Brewer BJ, Payen C, Raghuraman MK, Dunham MJ (2011) Origin-dependent inverted-repeat amplification: A replication-based model for generating palindromic amplicons. *PLoS Genet* 7:e1002016.
7. Shaw CJ, Lupski JR (2004) Implications of human genome architecture for rearrangement-based disorders: The genomic basis of disease. *Hum Mol Genet* 13: R57–R64.
8. Stratton MR (2011) Exploring the genomes of cancer cells: Progress and promise. *Science* 331:1553–1558.
9. Schimke RT (1984) Gene amplification in cultured animal cells. *Cell* 37:705–713.
10. Anderson RP, Roth JR (1977) Tandem genetic duplications in phage and bacteria. *Annu Rev Microbiol* 31:473–505.
11. Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM (2017) The dynamics of molecular evolution over 60,000 generations. *Nature* 551:45–50.
12. Gresham D, et al. (2010) Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the GAP1 locus. *Proc Natl Acad Sci USA* 107:18551–18556.
13. Iafrate AJ, et al. (2004) Detection of large-scale variation in the human genome. *Nat Genet* 36:949–951.
14. Sebat J, et al. (2004) Large-scale copy number polymorphism in the human genome. *Science* 305:525–528.
15. Flores M, et al. (2000) Prediction, identification, and artificial selection of DNA rearrangements in *Rhizobium*: Toward a natural genomic design. *Proc Natl Acad Sci USA* 97:9138–9143.
16. Mavingui P, et al. (2002) Dynamics of genome architecture in *Rhizobium* sp. strain NGR234. *J Bacteriol* 184:171–176.
17. Flores M, et al. (2007) Recurrent DNA inversion rearrangements in the human genome. *Proc Natl Acad Sci USA* 104:6099–6106.
18. Goffeau A, et al. (1996) Life with 6000 genes. *Science* 274:546, 563–7.
19. Engel SR, et al. (2014) The reference genome sequence of *Saccharomyces cerevisiae*: Then and now. *G3 (Bethesda)* 4:389–398.
20. Pennaneach V, Kolodner RD (2009) Stabilization of dicentric translocations through secondary rearrangements mediated by multiple mechanisms in *S. cerevisiae*. *PLoS One* 4:e6389.
21. Palacios-Flores K, et al. (2018) A perfect match genomic landscape provides a unified framework for the precise detection of variation in natural and synthetic haploid genomes. *Genetics* 208:1631–1641.
22. Chan JE, Kolodner RD (2011) A genetic and structural study of genome rearrangements mediated by high copy repeat Ty1 elements. *PLoS Genet* 7:e1002089.
23. Kupiec M, Petes TD (1988) Allelic and ectopic recombination between Ty elements in yeast. *Genetics* 119:549–559.
24. Dunham MJ, et al. (2002) Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 99:16144–16149.
25. Argueso JL, et al. (2008) Double-strand breaks associated with repetitive DNA can reshape the genome. *Proc Natl Acad Sci USA* 105:11845–11850.
26. Hou J, Friedrich A, de Montigny J, Schacherer J (2014) Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. *Curr Biol* 24:1153–1159.
27. Yue JX, et al. (2017) Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat Genet* 49:913–924.
28. Peter J, et al. (2018) Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556:339–344.
29. Shao Y, et al. (2018) Creating a functional single-chromosome yeast. *Nature* 560: 331–335.
30. Luo J, Sun X, Cormack BP, Boeke JD (2018) Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* 560:392–396.
31. Bailey JA, Liu G, Eichler EE (2003) An *Alu* transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* 73:823–834.
32. Deininger PL, Batzer MA (1999) *Alu* repeats and human disease. *Mol Genet Metab* 67: 183–193.
33. Delcher AL, Phillippy A, Carlton J, Salzberg SL (2002) Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* 30:2478–2483.
34. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359.