



**INSTITUTO POTOSINO DE INVESTIGACIÓN  
CIENTÍFICA Y TECNOLÓGICA, A.C.**

**POSGRADO EN CIENCIAS EN BIOLOGÍA MOLECULAR**

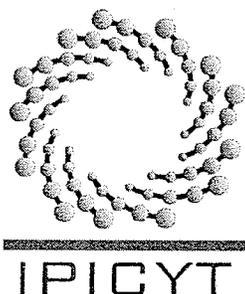
**Construcción y secuenciación de una colección de  
clonas de cromosomas bacterianos artificiales de una  
levadura cervecera**

Tesis que presenta  
**Cintia Noemí Gómez Muñoz**

Para obtener el grado de  
**Maestra en Ciencias en Biología Molecular**

**Codirectores de la Tesis:**  
**Dra. Lina Raquel Riego Ruiz**  
**Dr. Luis Cástulo Damas Buenrostro**

San Luis Potosí, S.L.P., 23 de enero de 2015



## Constancia de aprobación de la tesis

La tesis *“Construcción y secuenciación de una colección de clonas de cromosomas bacterianos artificiales de una levadura cervecera”* presentada para obtener el Grado de Maestra en Ciencias en Biología Molecular fue elaborada por Cintia Noemí Gómez Muñoz y aprobada el **veintitrés de enero del dos mil quince** por los suscritos, designados por el Colegio de Profesores de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C.

**Dra. Lina Raquel Riego Ruiz**  
Codirector de la tesis

**Dr. Luis Cástulo Damas Buenrostro**  
Codirector de la tesis

**Dra. Irene Beatriz Cástano Navarro**  
Miembro del Comité Tutorial

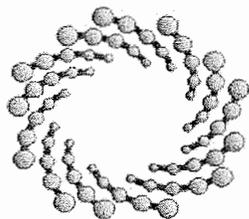


## **Créditos Institucionales**

Esta tesis fue elaborada en el Laboratorio de Genómica Funcional y Comparativa de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C., bajo la codirección los doctores Lina Raquel Riego Ruiz y Luis Cástulo Damas Buenrostro.

Este proyecto fue apoyado por la Convocatoria del Programa de Estímulos a la Investigación, Desarrollo Tecnológico e Innovación del Consejo Nacional de Ciencia y Tecnología (CONACYT) y la cervecería Cuauhtémoc Moctezuma S.A. de C.V.

Durante la realización del trabajo el autor recibió una beca académica del Consejo Nacional de Ciencia y Tecnología (No. 340255), del Instituto Potosino de Investigación Científica y Tecnológica, A. C. y una beca de recursos propios del proyecto S-2743 otorgado a la Dra. Lina Raquel Riego Ruiz.



**IPICYT**

# Instituto Potosino de Investigación Científica y Tecnológica, A.C.

## Acta de Examen de Grado

El Secretario Académico del Instituto Potosino de Investigación Científica y Tecnológica, A.C., certifica que en el Acta 126 del Libro Primero de Actas de Exámenes de Grado del Programa de Maestría en Ciencias en Biología Molecular está asentado lo siguiente:

En la ciudad de San Luis Potosí a los 23 días del mes de enero del año 2015, se reunió a las 13:00 horas en las instalaciones del Instituto Potosino de Investigación Científica y Tecnológica, A.C., el Jurado integrado por:

**Dr. Luis Cástulo Damas Buenrostro**  
**Dra. Irene Beatriz Castaño Navarro**  
**Dra. Lina Raquel Riego Ruiz**

**Presidente**  
**Secretaria**  
**Sinodal**

**CCM**  
**IPICYT**  
**IPICYT**

a fin de efectuar el examen, que para obtener el Grado de:

**MAESTRA EN CIENCIAS EN BIOLOGÍA MOLECULAR**

sustentó la C.

**Cintia Noemí Gómez Muñoz**

sobre la Tesis intitulada:

*Construcción y secuenciación de una colección de clonas de cromosomas bacterianos artificiales de una levadura cervecera*

que se desarrolló bajo la dirección de

**Dra. Lina Raquel Riego Ruiz**  
**Dr. Luis Cástulo Damas Buenrostro (CCM)**

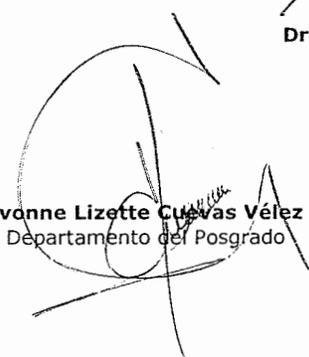
El Jurado, después de deliberar, determinó

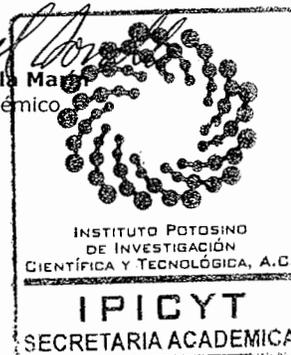
**APROBARLA**

Dándose por terminado el acto a las 14:20 horas, procediendo a la firma del Acta los integrantes del Jurado. Dando fe el Secretario Académico del Instituto.

A petición de la interesada y para los fines que a la misma convengan, se extiende el presente documento en la ciudad de San Luis Potosí, S.L.P., México, a los 23 días del mes de enero de 2015.

  
**Dr. Marcial Bonilla** Mar  
Secretario Académico

  
**Mtra. Ivonne Lizette Cuevas Vélez**  
Jefa del Departamento del Posgrado



A mi familia por su gran afecto. Gracias por permanecer unidos y apoyándonos a pesar de la distancia.

A los amigos que estuvieron a mi lado alentándome cuando más lo necesitaba.

## **Agradecimientos**

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo del Programa de Estímulos a la Investigación, Desarrollo Tecnológico e Innovación y la beca de estudios de maestría.

A mis codirectores de tesis, la Dra. Lina Raquel Riego Ruiz y el Dr. Luis Cástulo, y a la Dra. Irene Beatriz Castaño Navarro, por su confianza y apoyo a este proyecto.

A las personas que contribuyeron a la realización de este trabajo, el Dr. Nicolás Gómez Hernández y la Biól. Mireya Guadalupe Sánchez Garza por su apoyo técnico, y el M.C. Javier Israel Montalvo Arredondo por su asistencia en los análisis bioinformáticos.

Al Laboratorio Nacional de Biotecnología Agrícola, Médica y Ambiental (LANBAMA) por los servicios prestados.

## Contenido

Constancia de aprobación de la tesis	ii
Créditos institucionales	iii
Acta de examen	iv
Dedicatorias	v
Agradecimientos	vi
Lista de tablas	viii
Lista de figuras	ix
Anexos	x
Resumen	xi
Abstract	xii
Introducción	1
Materiales y Métodos	2
Resultados	6
Discusiones	12
Agradecimientos	16
Referencias	16
Tablas y Figuras	20
Anexos	36

## Lista de tablas

Tabla 1. Resultados de la transformación de prueba con insertos control	21
Tabla 2. Resumen de la caracterización de BACs con insertos experimentales	22
Tabla 3. Resultados de la secuenciación de los extremos del vector y análisis de las secuencias	23
Tabla 4. Tipos de pares de BES según sus parámetros de alineamiento	24
Tabla 5. Resultados de los parámetros de alineamiento de los pares de BES concordantes, opuesto e indeterminados	25
Tabla 6. Comparación de los tamaños de inserto bioinformáticos contra el tamaño de inserto estimado verificado experimentalmente de los distintos tipos de pares de BES	26

## Lista de figuras

Figura 1. Evaluación de la calidad del DNA obtenido usando dos protocolos de aislamiento	27
Figura 2. Digestiones parciales de prueba del DNA experimental	28
Figura 3. Digestión en masa y doble selección de tamaño de insertos experimentales	29
Figura 4. Caracterización con <i>NotI</i> de clonas aleatorias de la transformación experimental	30
Figura 5. Esquematizaciones de los distintos posibles alineamientos de los pares de BES	31
Figura 6. Validaciones de los alineamientos y estimación del tamaño de inserto de los distintos tipos de pares de BES	32
Figura 7. Clonas que unen scaffolds diferentes	33
Figura 8. Clonas que detectan ensamblajes incorrectos	34
Figura 9. Clona de par de BES opuesto	35

## Anexos

Anexo 1. Protocolo de electroforesis de inversión de campo para la separación de fragmentos entre 3 a 70 kb en el equipo Pippin Pulse de Sage Science	37
Anexo 2. Mapa y primers del vector pSMART BAC de Lucigen	38
Anexo 3. Detección de clonas sin inserto mediante PCR de colonia	39
Anexo 4. Diagrama de flujo del análisis de las secuencias de los extremos del vector	40
Anexo 5. Electrocariotipo de <i>Saccharomyces cerevisiae</i> y la levadura cervecera	41
Anexo 6. Lista de los pares de BES con alineamiento contra los scaffolds del ensamblaje de la levadura cervecera y el cromosoma de <i>Saccharomyces cerevisiae</i> o <i>Saccharomyces bayanus</i> contra el que alinean dichos scaffolds.	42
Anexo 7. Resumen rearrreglos propuestos y cromosomas de <i>Saccharomyces cerevisiae</i> o <i>Saccharomyces bayanus</i> contra el que alinean los scaffolds individuales de cada rearrreglo.	44

## Resumen

### **Construcción y secuenciación de una colección de clonas de cromosomas bacterianos artificiales de una levadura cervecera**

Las nuevas tecnologías de secuenciación permitieron la obtención de la secuencia genómica de una cepa de levadura cervecera tipo lager; sin embargo, el tamaño corto de las lecturas de estas tecnologías dificulta el proceso de ensamblaje. Una opción para mejorar el borrador de la secuencia es incorporar información de fuentes alternativas como lo son los mapas físicos basados en BACs. Por lo anterior, desarrollamos un protocolo para la construcción de clonas BACs con insertos de DNA de la levadura y obtuvimos las secuencias del extremo del vector (BES) de una colección de ellas. Encontramos que estos BESs pueden contribuir a la mejora del ensamblaje de la secuencia.

**PALABRAS CLAVE.** Nuevas Tecnologías de Secuenciación, ensamblaje del genoma, cromosomas bacterianos artificiales, secuencias de los extremos del vector.

## **Abstract**

### **Construction and sequencing of a bacterial artificial chromosome clone collection of a brewing yeast**

Next-generation sequencing technologies allowed the retrieval of the genome sequence of a lager beer yeast strain; however, the short read size of these technologies hinders the assembly process. One option to improve the draft sequence is by incorporating information from alternative sources such as BAC-based physical maps. Therefore, we developed a protocol for the construction of BAC clones with yeast DNA inserts and obtained the BAC-end sequences (BES) of a collection of them. We found that these BESs can contribute to the improvement of the sequence assembly.

**KEY WORDS.** Next-generation Sequencing Technologies, genome assembly, bacterial artificial chromosome, BAC-end sequencing.

## 1 **Introducción**

2 La cerveza más ampliamente distribuida en la industria cervecera es la tipo lager. Dicha  
3 cerveza es producida con una especie híbrida de levadura entre *Saccharomyces cerevisiae* y  
4 *Saccharomyces bayanus* o posiblemente *Saccharomyces eubayanus*, denominada  
5 *Saccharomyces pastorianus* (Martini y Martini, 1987; Libkind et al., 2011). La  
6 determinación de la secuencia genómica completa de la levadura lager es uno de los  
7 abordajes más prometedores para su explotación industrial (Nakao et al., 2009). Es por ello  
8 que la cervecería Cuauhtémoc Moctezuma se dio a la tarea de secuenciar el genoma de una  
9 cepa cervecera.

10 Las características de las nuevas tecnologías de secuenciación (o NGS por sus siglas en  
11 inglés), tales como su costo reducido y gran calibre, permitieron obtener en 2011 el primer  
12 borrador de la secuencia. Sin embargo, otras de las características de estas tecnologías,  
13 específicamente el tamaño corto de las lecturas dificulta el proceso de ensamblaje  
14 generando un genoma altamente fragmentado. Ya se han hecho estudios que demuestran  
15 que el uso exclusivo de lecturas cortas no es suficiente para el ensamblaje *de novo* (Alkan  
16 et al., 2011).

17 Se han adoptado diferentes estrategias para aumentar la calidad del ensamblaje del  
18 genoma las cuales incluyen re-secuenciación con diferentes plataformas e incorporación de  
19 la información de las bibliotecas de extremos pareados al ensamblaje (scaffolding)  
20 (Elizondo-González, 2014). Aun así, por la naturaleza de los datos y no tanto por las  
21 capacidades del programa, es posible que haya cientos y hasta miles de ensamblajes  
22 incorrectos que no se detectan (Salzberg y York, 2005). Más aún, debido a que el proceso  
23 de ensamblaje es meramente computacional, no hay manera de saber si este representa la  
24 realidad más que con la incorporación de evidencia experimental.

25 Por ello, es recomendable llevar una etapa de terminado y validación de la secuencia  
26 mediante la incorporación de información de fuentes alternativas. Una de las formas más  
27 rápidas para generar este tipo de datos es mediante la incorporación de mapas físicos al  
28 ensamblaje (Meyers et al., 2004). Actualmente, diversos grupos han reconocido el potencial  
29 de los mapas físicos para auxiliar y validar el proceso de ensamblaje de la secuencia. Por  
30 ejemplo, van Oeveren y colaboradores han desarrollado una tecnología denominada Whole

31 Genome Profiling (WGP) la cual ya se ha utilizado para asistir el ensamblaje *de novo* con  
32 NGS de distintos genomas de plantas (van Oeveren et al., 2011; Philippe et al., 2012).  
33 Recientemente, otro grupo ha utilizado información de mapas físicos para validar y mejorar  
34 ensamblajes con NGS (Pan et al., 2014; Dalloul et al., 2014). Con ello pudieron identificar  
35 cientos de regiones con discrepancias y extender el tamaño de los ensamblajes.

36 La mayoría de las técnicas de mapeo físico están basadas en la caracterización de  
37 bibliotecas de clonas de insertos grandes como los Cromosomas Bacterianos Artificiales (o  
38 BACs, por sus siglas en inglés) (Meyers et al., 2004; Shizuya et al., 1992). La  
39 incorporación de los mapas físicos al ensamblaje del genoma se realiza mediante las  
40 secuencias de los extremos del vector (o BES, por sus siglas en inglés) (Nelson y  
41 Soderlund, 2013).

42 La construcción de una biblioteca de BACs es conceptualmente sencilla pero  
43 técnicamente demandante. Es por ello que nos propusimos desarrollar un protocolo para la  
44 construcción de clonas BACs con insertos de DNA de la levadura cervecera.  
45 Posteriormente realizamos una prueba preliminar de secuenciación de los extremos del  
46 vector para valorar la utilidad de la estrategia del uso de clonas de insertos grandes para la  
47 validación y mejora del ensamblaje del genoma.

## 48 **Materiales y Métodos**

### 49 **Aislamiento de DNA genómico experimental y evaluación de su calidad**

50 Realizamos todos los cultivos de la levadura cervecera en medio YPD (extracto de levadura  
51 10 g/L, peptona de caseína 20g/L y dextrosa anhidra 20 g/L). Aislamos DNA genómico en  
52 solución con el protocolo de Hoffman y Winston (1987) modificado. Concentramos 5 mL  
53 de cultivo de la levadura en un volumen final de 1 mL centrifugando a 16,000×g por 3 min.  
54 Centrifugamos nuevamente y lisamos la pastilla con 200 µL de buffer de lisis (tritón X-100  
55 2%, SDS 1%, NaCl 100 mM, Tris-HCl 10 mM pH 8.0 y EDTA 1 mM pH 8.0), ~100 µL de  
56 perlas de vidrio (0.45-0.52 mm), 200 µL de fenol pH 8.0 a 4°C y agitación vigorosa por 3  
57 min. Añadimos 200 µL de buffer TE 1X y centrifugamos a >10,000×g por 1 min.  
58 Transferimos el sobrenadante, añadimos 3 µL de RNAsa 10 mg/mL e incubamos 20 min a  
59 37°C. Precipitamos con 1 mL de etanol al 96% frío (-20°C) y 1/10 volúmenes de acetato de

60 sodio 3M. Centrifugamos a 16,000×g por 5 min, lavamos con 1 mL de etanol al 70% frío (-  
61 20°C) y volvimos a centrifugar. Por último, resuspendimos la pastilla en TE 1X. La  
62 purificación del DNA en bloques de agarosa la realizamos con el kit CHEF Yeast Genomic  
63 DNA Plug de Bio-Rad siguiendo las instrucciones del fabricante. Evaluamos la calidad del  
64 DNA obtenido mediante electroforesis de inversión de campo (o FIGE, por sus siglas en  
65 inglés) con un protocolo pre-establecido para la separación de fragmentos entre 3 y 70 Kb:  
66 gel de agarosa para campo pulsado de Molecular Sigma Biology (agarosa-TBE 0.5X al 1%)  
67 en el equipo Pippin Pulse de Sage Science (75 V, 14 h; Anexo 1).

### 68 **Digestiones parciales de prueba del DNA experimental**

69 Seguimos el protocolo de Peterson et al. (2002) con algunas modificaciones. Disminuimos  
70 la concentración de EDTA de los bloques de DNA en agarosa con el buffer de lavado del  
71 kit de Bio-Rad a 0.1X por 1 h a temperatura ambiente y agitación suave, y dimos otro  
72 lavado rápido con esta misma solución. Equilibramos los bloques con el buffer de reacción  
73 CutSmart de New England Biolabs (CH<sub>3</sub>CO<sub>2</sub>K 50 mM, Tris-Acetato 20 mM,  
74 Mg(CH<sub>3</sub>COO)<sub>2</sub> 10 mM, BSA 100 µg/mL, pH 7.9) por 1 h a temperatura ambiente.  
75 Probamos ¼ de bloque macerado (~25µL) en 250 µL de buffer de reacción con 0, 0.2, 0.5,  
76 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 10.0 U de *Sau3AI* o *BamHI* así como con enzima concentrada  
77 (12.5 U y 20.0 U, respectivamente). Colocamos las reacciones 1 h en hielo y después las  
78 incubamos en un baño de agua a 37°C por 5 y 10 min para *Sau3AI* y por 15 y 20 min para  
79 *BamHI*. Detuvimos las reacciones con 30 µL de EDTA 0.5 M. Analizamos las digestiones  
80 mediante electroforesis de campo pulsado (o PFGE por sus siglas en inglés) con un  
81 protocolo para separar fragmentos de entre 50 a 300 Kb: gel de agarosa para campo pulsado  
82 de Molecular Sigma Biology (agarosa-TBE 0.5X al 1%) en el equipo CHEF Mapper XA de  
83 Bio-Rad (6.0 V/cm, 120°, 1-40 s, lineal, 18 h, 14°C).

### 84 **Ligación y electroporación de las muestras**

85 Ligamos 50 ng del vector pSMART BAC *BamHI* del CopyRight v2.0 BAC Cloning Kit de  
86 Lucigen que ya se encuentra digerido y desfosforilado (Anexo 2) con la cantidad indicada  
87 de insertos control o experimentales, 4 U de ligasa y buffer de ligación a 1X incluidos en el  
88 kit, en un volumen final de 100 µL. Incubamos las muestras a 16°C toda la noche (16 h).  
89 Posteriormente las microdializamos en un filtro de nitrocelulosa con un poro de 0.025 µm

90 sobre polietilenglicol al 10% estéril en una caja Petri durante 90 min. Preparamos en frío 20  
91  $\mu$ L de las células de *Escherichia coli* DH10B modificadas incluidas en el kit con 5  $\mu$ L de la  
92 muestra y las electroporamos utilizando el Gene Pulser II de Bio-Rad (20 kV/cm, 25  $\mu$ F y  
93 200 $\Omega$ ). Recuperamos las células en 1 mL de medio de recuperación incluido en el kit e  
94 incubamos a 37°C durante 1 h en agitación. Posteriormente sembramos en medio de  
95 selección YT Cm IPTG/X-Gal Sac (bacto-triptona 8 g/L, extracto de levadura 5 g/L, NaCl  
96 5 g/L, agar 15 g/L, cloranfenicol 12.5  $\mu$ g/mL, X-Gal 40  $\mu$ g/mL, IPTG 0.8 mM y sacarosa  
97 5%).

### 98 **Digestión en masa y doble selección de tamaño de insertos experimentales**

99 Digerimos 20 bloques de DNA macerados (~100  $\mu$ L cada uno) con 4 U/mL de *Sau3AI*  
100 cada uno en 1 mL de buffer de reacción. Incubamos a 37°C por 5 min y detuvimos la  
101 reacción con 150  $\mu$ L de EDTA 0.5 M. Posteriormente, separamos los bloques en un  
102 megacarril de 4  $\times$  0.5 cm mediante PFGE con un protocolo para separar fragmentos de  
103 entre 50 a 300 Kb: gel de agarosa para campo pulsado de Molecular Sigma Biology  
104 (agarosa-TBE 0.5X al 1%) en el equipo CHEF Mapper XA de Bio-Rad (6.0 V/cm, 120°, 1-  
105 40 s, lineal, 18 h, 14°C). Seleccionamos la región de entre 50 y 100 Kb sin exponerla a  
106 bromuro de etidio ni a luz UV y la dividimos en dos porciones (superior e inferior). Para  
107 eliminar fragmentos de menor tamaño que hayan sido arrastrados durante la primera  
108 selección, separamos nuevamente las porciones en un megacarril de 8  $\times$  0.5 cm con el  
109 mismo protocolo de PFGE y volvimos a seleccionar la región correspondiente.

### 110 **Electroelución y cuantificación de los insertos experimentales**

111 Electroeluímos el DNA seleccionado de la agarosa con el sistema Elutrap de Whatman,  
112 según las recomendaciones de los fabricantes (3.0 V/cm, 8 h, TAE 1X). Posteriormente  
113 visualizamos los insertos experimentales en un gel de electroforesis (agarosa-TAE 1X al  
114 1%) y calculamos su concentración comparando la intensidad de la banda contra un  
115 marcador de peso molecular conocido con el programa ImageJ (Rasband, 2012).

### 116 **Caracterización de clonas mediante *NotI* y estimación del tamaño de inserto**

117 Purificamos los vectores BAC siguiendo el protocolo de Farrar y Donnison (2007) o  
118 utilizando el Perfectprep BAC kit de 5 Prime. Las muestras se digirieron con *NotI* y se  
119 resolvieron mediante PFGE con un protocolo para separar fragmentos de entre 3 a 200 Kb:

120 gel de agarosa para campo pulsado de Molecular Sigma Biology (agarosa-TBE 0.5X al 1%)  
121 en el equipo CHEF Mapper XA de Bio-Rad (6.0 V/cm, 120°, 5-15 s, lineal, 16 h, 14°C).  
122 Estimamos el tamaño de inserto midiendo las distancias de los frentes de corrida de las  
123 bandas y comparándolas contra el marcador de peso molecular.

#### 124 **Detección de clonas sin inserto mediante PCR de colonia**

125 Hicimos la detección de clonas sin inserto a partir de una PCR de colonia usando el buffer  
126 Green Go Taq Reaction de Promega a 1X, MgCl<sub>2</sub> 1.5 mM, dNTPs 0.2 mM, primers  
127 forward (5'-TTG ACC ATG TTG GTA TGA TTT-3') y reverse (5'-CAG TCC AGT TAC  
128 GCT GGA GTC-3') (Anexo 2) 150 nM y 1 µL de *Taq* DNA polimerasa recombinante en  
129 un volumen final de 30 µL. El programa de PCR fue: desnaturalización a 94 °C por 5 min,  
130 30 ciclos de 94 °C por 45 s, 57°C por 30 s y 72°C por 60 s, y una extensión final a 72°C  
131 por 60 s. Analizamos los resultados en un gel de electroforesis (agarosa-TAE 1X al 2%)  
132 donde la presencia de una banda indica una clona sin inserto (Anexo 3).

#### 133 **Picado y conservación de la colección de BACs**

134 Cada clona se picó manualmente con un palillo estéril y se conservó en una copia maestra  
135 en placas de poliestireno de 96 pozos fondo "U" con 150 µL de medio de congelación  
136 (peptona de caseína 10 g/L, extracto de levadura 5 g/L, NaCl 10g/L, KH<sub>2</sub>PO<sub>4</sub> 13 mM,  
137 K<sub>2</sub>HPO<sub>4</sub> 36 mM, Na<sub>3</sub>C<sub>6</sub>H<sub>5</sub>O<sub>7</sub> 1.7 mM, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 6.8 mM, glicerol 4.4% v/v, MgSO<sub>4</sub> 0.4  
138 mM, y cloranfenicol 12.5 µg/mL). Las placas con las clonas se crecieron en agitación a  
139 37°C y 250 rpm por 24 h. Posteriormente, realizamos una segunda copia utilizando un  
140 replicador de placas y cultivando la réplica bajo las mismas condiciones usadas para la  
141 copia maestra. Las placas fueron cubiertas con sellos de aluminio y almacenadas a -80°C.

#### 142 **Secuenciación de los extremos del vector, análisis de las secuencias y alineamiento** 143 **contra el genoma de la levadura cervecera**

144 Los vectores se purificaron y secuenciaron por ambos extremos en el Clemson University  
145 Genomics Institute (Carolina del Sur, EE.UU.) con los primers forward (5'-TTG ACC  
146 ATG TTG GTA TGA TTT-3') y reverse (5'-CAG TCC AGT TAC GCT GGA GTC-3')  
147 (Anexo 2). Recortamos los extremos de las lecturas que tuvieran un valor de Phred Q<28 y  
148 agrupamos los pares de BES que pasaran este filtro de calidad. Alineamos los pares de BES  
149 de calidad contra la versión 3 del ensamblaje del genoma de la levadura cervecera

150 (Elizondo-González, 2014) utilizando BLAST+ con un umbral de similitud de valor de E =  
151  $1 \times 10^{-11}$  (Altschul et al., 1990; Camacho et al., 2009). Determinamos la orientación de los  
152 pares de BES según la referencia con un programa desarrollado en el laboratorio y escrito  
153 en lenguaje Python (Anexo 4). Los alineamientos contra la base de datos Nucleotide (nr) de  
154 los pares de BES falsos y sin alineamiento los realizamos en la página de internet del  
155 National Center for Biotechnology Information (NCBI) utilizando los parámetros pre-  
156 establecidos (Johnson et al., 2008).

## 157 **Resultados**

### 158 **Establecimiento protocolo de construcción de BACs**

#### 159 **Calidad del DNA experimental**

160 La clonación de fragmentos en el orden de las kilobases requiere partir de DNA genómico  
161 de Alto Peso Molecular (DNA APM). El DNA en solución, como el obtenido por los  
162 métodos convencionales de aislamiento de ácidos nucleicos, presenta rompimiento por  
163 estrés mecánico. Para evitar esto, el aislamiento de DNA APM se realiza embebido en  
164 bloques de agarosa.

165 Aislamos DNA genómico experimental en bloques de agarosa y evaluamos su calidad  
166 comparándolo con DNA obtenido mediante un protocolo en solución. Observamos que el  
167 DNA obtenido con el protocolo en solución se encontró alrededor de los 35 Kb mientras  
168 que el DNA en bloques de agarosa fue de mayor tamaño que la banda superior del  
169 marcador  $\lambda$  que es de 48.5 Kb (Figura 1).

#### 170 **Condiciones óptimas para la preparación de insertos experimentales**

171 El siguiente paso después de la obtención del DNA APM, es la preparación de los  
172 insertos experimentales de las clonas. Esto se realiza mediante una digestión parcial en la  
173 que no todos los sitios de reconocimiento de la enzima son cortados. Esto se logra  
174 modificando las condiciones óptimas de corte de la enzima.

175 Realizamos varias digestiones parciales de prueba con el DNA experimental utilizando  
176 las enzimas de restricción *Sau3AI* y *BamHI* a diferentes concentraciones y tiempos de  
177 incubación. Observamos que el tamaño del DNA disminuyó conforme aumentó la

178 concentración de enzima en la reacción (Figura 2). No apreciamos una diferencia clara  
179 entre los diferentes tiempos de incubación probados para las dos enzimas pero establecimos  
180 que las mejores condiciones para obtener la mayor proporción de fragmentos dentro del  
181 intervalo deseado de 50 a 100 kb fueron usando 1.0 U de enzima durante 5 min de  
182 incubación para *Sau3AI* (Figura 2A) y 2.0 U de enzima durante 15 min de incubación para  
183 *BamHI* (Figura 2B).

#### 184 **Transformación de prueba con insertos control**

185 Realizamos una transformación de prueba con una ligación usando 160 ng del inserto  
186 control y 50 ng del vector, con sus respectivos controles (control negativo de la  
187 transformación, control negativo de la ligación y control de la calidad del vector) para  
188 determinar el número de transformantes promedio posibles por evento de electroporación  
189 (ufc/mL) con las condiciones de ligación y electroporación previamente establecidas.

190 Obtuvimos un promedio de 16,800 transformantes por evento de electroporación de las  
191 células del kit con la reacción de ligación con el inserto control en medio de selección. No  
192 creció ninguna clona en las placas sembradas con los controles en el mismo medio (Tabla  
193 1). Caracterizamos con *NotI* algunas clonas de esta transformación y determinamos la  
194 presencia del inserto control en por lo menos 80% de éstas (n = 20) (datos no mostrados).

#### 195 **Construcción de BACs con insertos experimentales**

##### 196 **Preparación de insertos experimentales**

197 Una vez optimizada la transformación de prueba, el siguiente paso consistió en repetir el  
198 procedimiento pero esta vez utilizando el DNA experimental. Realizamos una digestión en  
199 masa del DNA APM de la levadura e hicimos una doble selección de tamaño de los  
200 insertos. Observamos un barrido en la primera selección de tamaño producto de la digestión  
201 parcial del DNA (Figura 3A). En la segunda selección observamos un ligero barrido en la  
202 parte inferior de la región seleccionada (Figura 3B). La región sustraída apareció en negro y  
203 se encontró entre los 50 y 100 kb en ambos casos (Figura 3 A y B). Posteriormente  
204 electroeluímos de la agarosa los insertos experimentales de la segunda selección de tamaño,  
205 los cuantificamos y determinamos una concentración de 0.99 ng/μL.

206 **Transformación con insertos experimentales y caracterización de clonas**

207 Preparamos una reacción de ligación usando 73.26 ng del DNA electroeluido con 50 ng del  
208 vector y la transformamos con el protocolo estandarizado. Obtuvimos un total de 192  
209 transformantes por mililitro de medio recuperado (Tabla 2).

210 Para determinar la proporción de clonas con inserto, muestreamos y caracterizamos con  
211 *NotI* clonas aleatorias de esta transformación y determinamos una proporción de 7/20 (35%)  
212 clonas con inserto sobre el total de clonas. Estimamos el tamaño de los insertos y  
213 obtuvimos un promedio de 35.97±9.99 Kb (Figura 4 y Tabla 2).

214 **Secuenciación y análisis de los extremos del vector**

215 **Secuenciación de los extremos del vector y análisis de las secuencias**

216 Enviamos a secuenciar 96 clonas por ambos extremos y de estas 59 pares de secuencias  
217 (61.5%) pasaron los filtros de calidad. Obtuvimos un total de 91,670 bases secuenciadas  
218 con un tamaño promedio de las lecturas de 776.86±95.04 pb (Tabla 3).

219 **Alineamiento de los BES contra el genoma de la levadura cervecera**

220 Alineamos los 59 pares de BES de calidad contra la versión 3 del ensamblaje del genoma  
221 de la levadura cervecera que tiene una cobertura aproximada de 70X, un tamaño del  
222 genoma estimado de 22.7 Mb y que consta de 133 scaffolds (Elizondo-González, 2014). De  
223 los 59 pares de BES de calidad ninguno alineó más de una vez, 42 alinearon solo una y 17  
224 no alinearon en ninguna ocasión (Tabla 4).

225 Clasificamos los BES de calidad dependiendo de la dirección 5'→3' entre ellos en:  
226 concordantes si son convergentes (Figura 5A), opuestos si son divergentes (Figura 5B) e  
227 indeterminados si su dirección no se puede conocer debido a que alinean en scaffolds  
228 diferentes (Figura 5E). Encontramos que 37 pares de BES fueron concordantes, 1 fue  
229 opuesto y 4 indeterminados (Tabla 4). Los pares de BES que alinean en la misma cadena y  
230 en la misma dirección (5'→3' o positivos y 3'→5' o negativos) también son posibles; sin  
231 embargo, no encontramos algún caso de estos (Figura 5 C y D).

232 Observamos que los pares de BES concordantes se agrupaban en tres categorías  
233 dependiendo de sus parámetros de alineamiento. En el primer grupo consideramos que 21

234 de ellos eran pares de BES verdaderos ya que sus valores de longitud promedio de  
235 alineamiento ( $732.90 \pm 127.21$  pb), de bit score promedio ( $1,408.33 \pm 269.46$  pb) y tamaño de  
236 inserto bioinformático promedio ( $38,460.14 \pm 16,760.47$  pb) fueron altos con respecto a los  
237 otros grupos. En un segundo grupo encontramos 2 pares de BES intermedios dado que sus  
238 valores de longitud promedio de alineamiento ( $748.75 \pm 89.49$  pb) y de bit score promedio  
239 fueron altos ( $1,411.50 \pm 161.33$  pb), pero su tamaño de inserto bioinformático promedio fue  
240 pequeño ( $7,562.50$  pb). Por último consideramos que 14 pares de BES fueron falsos ya que  
241 sus valores de longitud promedio de alineamiento ( $121.36 \pm 73.98$  pb), de bit score promedio  
242 ( $236.68 \pm 146.58$  pb) y tamaño del inserto bioinformático promedio fueron bajos  
243 ( $137.29 \pm 73.75$  pb) con respecto a las categorías anteriores. La longitud de alineamiento del  
244 par de BES opuesto (796 pb) y la longitud promedio de alineamiento de los pares de BES  
245 indeterminados ( $675.5 \pm 98.09$  pb) fueron altos. Lo mismo sucedió con el valor de bit score  
246 del par de BES opuesto y el promedio de bit score de los pares de BES indeterminados  
247 ( $1557$  y  $1250.25 \pm 385.05$ , respectivamente); sin embargo, no pudimos determinar su tamaño  
248 de inserto bioinformático (Tabla 4 y Tabla 5).

#### 249 **Validaciones de los alineamientos y rearrreglos propuestos**

250 Retamos los pares de BES falsos y sin alineamiento contra la base de datos “Nucleotide”  
251 del NCBI y encontramos que estos alinean principalmente con la secuencia del vector  
252 pSMART (datos no mostrados).

253 Posteriormente caracterizamos con *NotI* algunas clonas seleccionadas al azar de los  
254 distintos tipos de pares de BES y estimamos su tamaño del inserto para validar los  
255 resultados del alineamiento. Encontramos que todas las clonas de pares de BES verdaderos  
256 ( $n = 5$ ) e intermedios ( $n = 2$ ) tuvieron inserto y además el tamaño de inserto bioinformático  
257 coincidió con el tamaño de inserto estimado. Así también comprobamos que los pares de  
258 BES falsos ( $n = 3$ ) y sin alineamiento ( $n = 3$ ) no tuvieron inserto. Por último, obtuvimos el  
259 tamaño estimado del inserto de los pares de BES indeterminados ( $n = 4$ ) y el par de BES  
260 opuesto ( $n = 1$ ) (Figura 6 y Tabla 6).

261 Determinamos la proporción de clonas con inserto a partir de los resultados de la  
262 secuenciación de los extremos de vector y encontramos que 28 de los 59 pares de BES  
263 (47.46%) tenían inserto.

264 También determinamos la proporción de clonas con inserto de las mismas clonas de los  
265 pares de BES de calidad mediante la prueba de PCR de colonia y encontramos que 29 de  
266 las 59 clonas tenían inserto (49.15%). Estos resultados coincidieron con los datos de la  
267 secuenciación de los extremos del vector en 58 de las 59 clonas (datos no mostrados).

268 Analizamos cada uno de los casos de pares de BES indeterminados y el caso de par de  
269 BES opuesto y propusimos un nuevo ensamblaje para cada uno. Para este propósito  
270 determinamos el tamaño de inserto hipotético de las clonas sumando las distancias de los  
271 extremos de dónde alinean los pares de BES hasta el final del scaffold. De los cuatro casos  
272 de pares de BES indeterminados dos unieron scaffolds (Figura 7) mientras que dos  
273 detectaron posibles ensamblajes incorrectos (Figura 8).

274 En el primer caso de clonas que unen scaffolds, un BES de la clona B04 alineó en el  
275 scaffold 38 en orientación negativa ( $3' \rightarrow 5'$ ) y el otro en el scaffold 64 en orientación  
276 positiva ( $5' \rightarrow 3'$ ). El tamaño del inserto hipotético fue de 29,121 pb mientras que el inserto  
277 estimado fue de 28,958 pb por lo que la diferencia entre ambos fue de 163 pb. Proponemos  
278 un rearrreglo en el que el scaffold 64 y el scaffold 38 están unidos en orientación positiva  
279 casi sin ningún espacio (Figura 7A).

280 La clona F08 es otro caso de clonas que unen scaffolds. Aquí un BES alinea en el  
281 scaffold 2 en orientación positiva y el otro en el scaffold 53 en orientación positiva. El  
282 tamaño del inserto hipotético (10,820 pb) es menor al tamaño de inserto estimado (20,988  
283 pb) por lo que el rearrreglo propuesto consta del scaffold 2 en orientación positiva, un  
284 espacio de 10,168 pb y el scaffold 53 en orientación negativa (Figura 7B).

285 En el caso de clonas que detectan posibles ensamblajes incorrectos, un BES de la clona  
286 G04 alineó en el scaffold 32 y otro en el scaffold 7; sin embargo, el tamaño de inserto  
287 hipotético (139,837 pb) fue mucho más grande que el tamaño de inserto estimado (36,444  
288 pb). Un tamaño de inserto tan grande es poco probable ya que seleccionamos insertos en un  
289 intervalo de 50 a 100 Kb. Por lo anterior proponemos un rearrreglo en el que el scaffold 32  
290 está fragmentado en dos porciones (scaffold 32-A y scaffold 32-B) en algún punto  
291 desconocido; sin embargo, no tenemos evidencia de que esto suceda para el scaffold 7. El  
292 rearrreglo propuesto consta del scaffold 32-A en orientación positiva, un espacio de un

293 tamaño desconocido y el scaffold 7 en orientación negativa más la porción del scaffold 32-  
294 B de un tamaño desconocido por separado (Figura 8A).

295 El caso de la clona H07 es otro ejemplo de pares de BES que detectan un posible  
296 ensamblaje incorrecto. En esta un BES alinea en el scaffold 1 y el otro en el scaffold 62; sin  
297 embargo, el tamaño de inserto hipotético (623,593 pb) es mucho más grande que el tamaño  
298 de inserto estimado (34,102 pb). Con lo anterior propusimos que el scaffold 1 debería estar  
299 fragmentado en dos porciones (scaffold 1-A y scaffold 1-B) en algún punto desconocido  
300 pero no encontramos evidencia de que así suceda para el scaffold 62. El rearrreglo propuesto  
301 constaría del scaffold 1-A en orientación positiva, un espacio de tamaño desconocido y el  
302 scaffold 62 en orientación negativa más la porción del scaffold 1-B de un tamaño  
303 desconocido por separado (Figura 8B).

304 En el caso de la clona F06 ambos BES alinean en el scaffold 10 pero en direcciones  
305 opuestas. No determinamos un tamaño de inserto hipotético pero el tamaño de inserto  
306 estimado de la clona fue de 25,674 pb. Posiblemente el scaffold 10 está unido por una  
307 región que debería estar por separado. El rearrreglo propuesto consta del scaffold 10-A de  
308 un tamaño desconocido, una porción del scaffold 10-B fragmentado y reorientado, y la  
309 porción que previamente unía el scaffold por separado (región 558,678-562,776 pb) (Figura  
310 9).

311 Alineamos los pares de BES de calidad contra dos versiones reportadas del genoma de *S.*  
312 *pastorianus* Weihenstephan 34/70 (Nakao et al., 2009, cobertura 6X; Walther et al., 2014,  
313 cobertura 18X). En general encontramos que había un número menor de pares de BES  
314 concordantes y un número mayor de pares de BES indeterminados. Esto quiere decir que  
315 algunos de los pares de BES que fueron concordantes cuando se compararon contra el  
316 ensamblaje de la levadura cervecera fueron indeterminados cuando se compararon con los  
317 ensamblajes de *S. pastorianus* Weihenstephan 34/70. Ninguno de los pares de BES  
318 indeterminados de las clonas B04, F08, G04 y H07 ni el par de BES opuesto de la clona  
319 F06 alineó en un mismo contig de las dos versiones por lo que no encontramos evidencia  
320 que sustentara los ensamblajes propuestos (datos no mostrados).

## 321 **Discusión**

### 322 **Construcción y calidad de la colección de BACs con insertos experimentales**

323 Para la construcción de la colección de BACs escogimos un intervalo de inserto de entre 50  
324 y 100 Kb, esto para tener bien representado el cromosoma más pequeño de *S. pastorianus*  
325 que es de ~200Kb (Anexo 5). Por ello, fue necesario encontrar un protocolo de aislamiento  
326 de DNA que produzca fragmentos mayores a 100 Kb. El DNA obtenido con el protocolo de  
327 aislamiento de DNA en bloques de agarosa es de mayor tamaño que el obtenido por el  
328 protocolo de DNA en solución y a su vez es más grande que la banda superior del marcador  
329 de fago lambda de 48.5 Kb. Por lo anterior, podemos decir que el DNA obtenido en  
330 bloques de agarosa es apto para su utilización en la preparación de insertos experimentales.

331 Es posible obtener condiciones óptimas para la preparación de insertos experimentales  
332 con *Sau3AI* y *BamHI*. Escogimos el menor tiempo de incubación para favorecer la  
333 generación de fragmentos más grandes. Finalmente, elegimos a *Sau3AI* ya que tiene un  
334 sitio de reconocimiento más pequeño (4 pb) que *BamHI* (6 pb), lo cual genera fragmentos  
335 más variados y disminuye el sesgo de la biblioteca (Seed et al., 1982).

336 La transformación de prueba con insertos control tuvo un alto número de transformantes  
337 por evento de electroporación; sin embargo, con insertos experimentales el número de  
338 transformantes que obtuvimos fue muy bajo. Esto podría deberse a que la transformación se  
339 optimizó utilizando insertos control que tienen un tamaño relativamente pequeño de  
340 alrededor de 15 Kb (datos no mostrados). En teoría, vectores más grandes como los  
341 generados en la ligación con insertos experimentales serían más difíciles de introducir a las  
342 células por electroporación en comparación con vectores más pequeños como los que  
343 tienen el inserto control (Sheng et al., 1995). Una opción para mejorar el número de  
344 transformantes obtenidas es optimizar el protocolo de electroporación utilizando BACs con  
345 insertos más grandes.

346 La proporción de clonas con insertos experimentales fue muy baja en comparación con  
347 las clonas con insertos control (35% contra 80%, respectivamente). Así también, la  
348 proporción de clonas con inserto experimentales determinado a partir de la secuenciación  
349 de los extremos del vector fue un poco mayor al determinado mediante caracterización con

350 *NotI* (47.46% contra 35%, respectivamente). Esto podría deberse al diferente número de  
351 clonas muestreadas.

352 Por otro lado, podemos decir que la prueba de PCR de colonia para detectar clonas sin  
353 inserto es útil ya que determinó correctamente la presencia o ausencia de inserto en 58 de  
354 59 clonas. Como muchas de las clonas secuenciadas no tienen insertos, podríamos utilizar  
355 la prueba de PCR de colonia para realizar un escrutinio rápido de las clonas antes de  
356 secuenciar.

357 El bajo número de transformantes obtenido en la transformación con insertos  
358 experimentales, así como la baja proporción de clonas con inserto pueden deberse a la baja  
359 concentración de insertos experimentales alcanzada con la electroelución (0.99 ng/ $\mu$ L). Es  
360 recomendable alcanzar como mínimo una relación inserto-vector de 5:1 (Peterson et al.,  
361 2002). Suponiendo un tamaño promedio de inserto de 75 Kb, deberíamos haber utilizado  
362 una cantidad total de 2500 ng contra 50 ng del vector; sin embargo, nosotros solo  
363 alcanzamos una cantidad de 73.26 ng totales en la ligación.

364 El tamaño promedio de los insertos experimentales coincidió con el tamaño  
365 bioinformático promedio determinado. El tamaño promedio de inserto experimental es  
366 menor al del intervalo seleccionado (50-100 Kb). Además, a pesar de la doble selección de  
367 tamaño encontramos clonas con insertos de alrededor de 8 Kb. Esto podría explicarse en  
368 parte a que los insertos pequeños tienen mayor probabilidad de inclusión que los grandes  
369 (Sheng et al., 1995). Las opciones para mejorar el tamaño promedio del inserto de la  
370 colección de BACs son seleccionar de 20 a 50 Kb más arriba del intervalo deseado (Farrar  
371 y Donnison, 2007) o utilizar un protocolo diferente de selección de tamaño (Osoegawa et  
372 al., 1998).

### 373 **Rearreglos propuestos y discrepancias con el ensamblaje**

374 La tasa de éxito de la secuenciación de los extremos del vector es muy baja, ya que solo  
375 61.5% de los pares de BES pasaron los filtros de calidad. Esto puede deberse a una baja  
376 eficiencia de la reacción de secuenciación de alguno de los oligonucleótidos y a que se  
377 necesita que ambas secuencias cumplan los parámetros establecidos. A pesar de esto, el  
378 tamaño promedio de las lecturas exitosas después del recorte es bueno.

379 Ningún par de BES alinea en más de una ocasión en el ensamblaje de la secuencia, lo  
380 cual podría deberse al bajo número de BES utilizados en el análisis.

381 Los pares de BES falsos y sin alineamiento que son lecturas del vector sin inserto  
382 pueden ser identificados y eliminados incorporando en la metodología un paso de  
383 alineamiento de los pares de BES contra el vector antes del alineamiento contra el  
384 ensamblaje del genoma.

385 Los pares de BES verdaderos e intermedios afianzan el ensamblaje ya que el tamaño de  
386 inserto bioinformático y el tamaño de inserto estimado de las clonas muestreadas  
387 coincidieron. Las pequeñas diferencias entre ellos pueden deberse al método de estimación  
388 del tamaño de los insertos en el gel o en su defecto ser reales. Esto podría confirmarse con  
389 una secuenciación profunda de las clonas.

390 Dado que la mayoría de los pares de BES fueron verdaderos podemos decir que el  
391 ensamblaje en general es de buena calidad. Los parámetros que se pudieron determinar de  
392 los pares de BES indeterminados fueron similares a los de los pares de BES verdaderos, por  
393 lo que podemos pensar que se tratan de alineamientos verdaderos. Dos pares de BES  
394 ayudaron a extender el ensamblaje, ya que unen scaffolds diferentes. En cambio, otros tres  
395 pares de BES tienen discrepancias con el ensamblaje lo que sugiere posibles ensamblajes  
396 incorrectos como lo reportado en otros trabajos (Pan et al., 2014). Los espacios y errores en  
397 el ensamblaje pueden atribuirse a la presencia de repetidos (Schatz et al., 2010), la falta de  
398 regiones difíciles de secuenciar (Kieleczawa et al., 2006) o al algoritmo utilizado para  
399 ensamblar los scaffolds (Salzberg y York, 2005). No obstante, no podemos descartar  
400 eventos de rearrreglos del inserto del plásmido in vivo (Jones et al., 1982).

401 El hecho de que no encontráramos evidencia que sustentara las nuevas propuestas del  
402 ensamblaje en los dos ensamblajes reportados de *S. pastorianus* Weihenstephan 34/70  
403 (Nakao et al. 2009; Walther et al., 2014) puede deberse a que tienen menor cobertura o a  
404 diferencias específicas de las cepas.

405 No obstante, cuando comparamos nuestros resultados con un trabajo previo en el que se  
406 alineó el ensamblaje de la levadura cervecera contra los genomas de *S. cerevisiae* y *S.*

407 *bayanus* (Elizondo-González, 2014) encontramos que los scaffolds unidos por los pares de  
408 BES alinean en el mismo cromosoma. Por ejemplo, los scaffolds 38 y 64 unidos por el par  
409 de BES de la clona B04 alinean con el cromosoma 16 de *S. cerevisiae*. Esto mismo ocurre  
410 con los scaffolds 2 y 53 unidos por el par de BES de la clona F08 los cuales alinean con el  
411 cromosoma 15 de *S. cerevisiae*. Lo anterior es otra evidencia que sustenta que los  
412 rearrreglos propuestos para estos casos son correctos. En cambio, los scaffolds que son  
413 fragmentados y rearrreglados alinean con distintos cromosomas y en ocasiones múltiples  
414 veces. Por ejemplo, el scaffold 32 separado por el par de BES de la clona G04 alinea con el  
415 cromosoma 3 y 5 de *S. cerevisiae*, mientras que el scaffold 7 alinea con el cromosoma 8 de  
416 *S. cerevisiae*. Esto podría representar una evidencia de que el scaffold 7 también deba ser  
417 separado. El scaffold 1 separado por el par de BES de la clona H07 alinea con el  
418 cromosoma 1 y 12 de *S. bayanus*, mientras que el scaffold 62 alinea también con el  
419 cromosoma 1 de *S. bayanus*. Esto podría confirmar que el scaffold 1-A y el scaffold 62  
420 forman parte de una misma región del cromosoma 1 de *S. bayanus* mientras que el scaffold  
421 1-B que separamos podría formar parte del cromosoma 12. El scaffold 10 separado por el  
422 par de BES de la clona F06 alinea con el cromosoma 10 de *S. cerevisiae* y el cromosoma 10  
423 de *S. bayanus*. Puede ser que estos cromosomas en el genoma de la levadura cervecera  
424 están ensamblados juntos debido a que son muy similares. Estos ejemplos reafirman que las  
425 clonas G04, H07 y F06 detectan ensamblajes incorrectos (Anexo 6 y 7).

#### 426 **Utilidad de la colección y validez de los resultados**

427 Establecimos un protocolo de construcción de clonas con insertos de DNA de la levadura  
428 cervecera; sin embargo, la calidad de la colección generada es baja. Estos resultados se  
429 deben a cuestiones técnicas que pueden ser mejoradas analizando los pasos clave de la  
430 metodología.

431 La construcción de una biblioteca de insertos grandes supone un esfuerzo adicional; sin  
432 embargo, este estudio demuestra que puede ser útil para mejorar la calidad del ensamblaje e  
433 incluso superior a las bibliotecas de extremos pareados utilizadas para secuenciación con  
434 Illumina con insertos de 350 y 8,000 pb (Elizondo-González, 2014).

435 A partir de los resultados del análisis de 96 muestras estamos proponiendo rearrreglos  
436 que posiblemente mejoren el ensamblaje. El alineamiento de los pares de BES contra el

437 genoma de la levadura cervecera valida la mayor parte del ensamblaje, une scaffolds y  
438 detecta posibles ensamblajes incorrectos. No obstante, aún es necesario demostrar que los  
439 rearrreglos propuestos son correctos; por ejemplo, secuenciando los espacios, determinando  
440 los puntos de quiebre de los scaffolds o encontrando un mayor número de clonas que  
441 alineen en esas regiones. De ser así este análisis demostraría que a pesar de tener una alta  
442 cobertura, el proceso de ensamblaje no está libre de errores y que es fundamental  
443 incorporar información de más de un tipo.

444 Nuestras perspectivas son optimizar el protocolo de la construcción de los BACs y  
445 construir la biblioteca con un número suficiente de clonas para tener representado el  
446 genoma de la levadura de 3 a 5 veces. Así también, uno de los objetivos a futuro es generar  
447 un mapa físico a partir de esta biblioteca e integrarlo al ensamblaje actual de la secuencia  
448 de la levadura cervecera.

#### 449 **Agradecimientos**

450 Este proyecto fue apoyado por la Convocatoria del Programa de Estímulos a la  
451 Investigación, Desarrollo Tecnológico e Innovación del Consejo Nacional de Ciencia y  
452 Tecnología (CONACYT). Durante su desarrollo Cintia Gómez-Muñoz recibió una beca  
453 para estudios de maestría No. 340255 por parte del CONACYT y una beca de recursos  
454 propios del proyecto S-2743 otorgado a la Dra. Lina Riego-Ruiz. Así también,  
455 agradecemos a la Cervecería Cuauhtémoc Moctezuma por el apoyo brindado a la  
456 realización del presente proyecto.

#### 457 **Referencias**

- 458 Alkan, C., Sajjadian, S., y Eichler, E.E. (2011). Limitations of next-generation genome  
459 sequence assembly. *Nat. Methods* 8, 61–65.
- 460 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., y Lipman, D.J. (1990). Basic local  
461 alignment search tool. *J. Mol. Biol.* 215, 403–410.
- 462 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., y Madden,  
463 T.L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.

464 Clarke, L., y Carbon, J. (1976). A colony bank containing synthetic Col El hybrid plasmids  
465 representative of the entire *E. coli* genome. *Cell* 9, 91–99.

466 Dalloul, R.A., Zimin, A.V., Settlage, R.E., Kim, S., y Reed, K.M. (2014). Next-generation  
467 sequencing strategies for characterizing the turkey genome. *Poult. Sci.* 93, 479–484.

468 Elizondo-González R. (2014). Ensamblaje del genoma de una levadura cervecera tipo lager  
469 (*Saccharomyces cerevisiae* var. *uvarum*). Tesis de maestría. Universidad Autónoma de  
470 Nuevo León.

471 Farrar, K., y Donnison, I.S. (2007). Construction and screening of BAC libraries made  
472 from *Brachypodium* genomic DNA. *Nat. Protocols* 2, 1661–1674.

473 Hoffman, C.S., y Winston, F. (1987). A ten-minute DNA preparation from yeast efficiently  
474 releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* 57, 267–272.

475 Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., y Madden, T.L.  
476 (2008). NCBI BLAST: a better web interface. *Nucleic. Acids. Res.* 36, W5–W9.

477 Jones, I.M., Primrose, S.B., y Ehrlich, S.D. (1982). Recombination between short direct  
478 repeats in a *recA* host. *Mol. Gen. Genet.* 188, 486–489.

479 Kieleczawa, J. (2006). Fundamentals of Sequencing of Difficult Templates—An Overview.  
480 *J. Biomol. Tech.* 17, 207–217.

481 Libkind, D., Hittinger, C.T., Valério, E., Gonçalves, C., Dover, J., Johnston, M.,  
482 Gonçalves, P., y Sampaio, J.P. (2011). Microbe domestication and the identification of  
483 the wild genetic stock of lager-brewing yeast. *Proc. Natl. Acad. Sci. U S A* 108, 14539–  
484 14544.

485 Martini, A.V., y Martini, A. (1987). Three newly delimited species of *Saccharomyces sensu*  
486 *stricto*. *Antonie Van Leeuwenhoek* 53, 77–84.

487 Meyers, B.C., Scalabrin, S., y Morgante, M. (2004). Mapping and sequencing complex  
488 genomes: let's get physical! *Nat. Rev. Genet.* 5, 578–588.

489 Nakao, Y., Kanamori, T., Itoh, T., Kodama, Y., Rainieri, S., Nakamura, N., Shimonaga, T.,  
490 Hattori, M., y Ashikari, T. (2009). Genome Sequence of the Lager Brewing Yeast, an  
491 Interspecies Hybrid. *DNA Res.* 16, 115–129.

492 Nelson, W.M., y Soderlund, C. (2013). Integrating sequence with FPC fingerprint maps. *J.*  
493 *Pharm. Sci.* 102, 469–479.

494 Van Oeveren, J., de Rooter, M., Jesse, T., van der Poel, H., Tang, J., Yalcin, F., Janssen, A.,  
495 Volpin, H., Stormo, K.E., Bogden, R., van Eijik M.J.T. y Prince, M (2011). Sequence-  
496 based physical mapping of complex genomes by whole genome profiling. *Genome Res.*  
497 21, 618–625.

498 Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J., y de Jong,  
499 P.J. (1998). An Improved Approach for Construction of Bacterial Artificial  
500 Chromosome Libraries. *Genomics* 52, 1–8.

501 Peterson, D.G., Tomkins, J.P., Frisch, D.A., Wing, R.A., y Paterson, A.H. (2002).  
502 Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated  
503 guide. *J. Agric. Genomics* 5, 1–100.

504 Philippe, R., Choulet, F., Paux, E., van Oeveren, J., Tang, J., Wittenberg, A.H.J., Janssen,  
505 A., van Eijk, M.J.T., Stormo, K., Alberti, A., Wincker, P., Akhunov, E., van der Vossen,  
506 E. y Feuillet C. (2012). Whole Genome Profiling provides a robust framework for  
507 physical mapping and sequencing in the highly complex and repetitive wheat genome.  
508 *BMC Genomics* 13, 47.

509 Rasband, W.S. (2012). ImageJ: Image processing and analysis in Java. *ASCL* -1, 06013.

510 Salzberg, S.L., y Yorke, J.A. (2005). Beware of mis-assembled genomes. *Bioinformatics*  
511 21, 4320–4321.

512 Schatz, M.C., Delcher, A.L., y Salzberg, S.L. (2010). Assembly of large genomes using  
513 second-generation sequencing. *Genome Res.* 20, 1165–1173.

- 514 Seed, B., Parker, R.C., y Davidson, N. (1982). Representation of DNA sequences in  
515 recombinant DNA libraries prepared by restriction enzyme partial digestion. *Gene* 19,  
516 201–209.
- 517 Sheng, Y., Mancino, V., y Birren, B. (1995). Transformation of *Escherichia coli* with large  
518 DNA molecules by electroporation. *Nucleic Acids Res.* 23, 1990–1996.
- 519 Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., y Simon, M.  
520 (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA  
521 in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. U S A* 89,  
522 8794–8797.
- 523 Walther, A., Hesselbart, A., y Wendland, J. (2014). Genome Sequence of *Saccharomyces*  
524 *carlsbergensis*, the World’s First Pure Culture Lager Yeast. *G3 (Bethesda)* 4, 783–793.

## **Tablas y Figuras**

**Tabla 1. Resultados de la transformación de prueba con insertos control.**

<b>Muestra</b>	<b>Contenido</b>	<b>Promedio transformantes<sup>a</sup></b>
Transformación insertos control	Células + VIL	16,800 ufc, n=3
Control negativo transformación	Células + agua Milli-Q	0 ufc, n=1
Control negativo ligación	Células + VI	0 ufc, n=1
Control calidad vector	Células + VL	0 ufc, n=1

a. Número de transformantes por evento de electroporación de las células del kit con las distintas muestras en medio de selección (YT Cm IPTG/X-Gal Sac).

V, vector; I, inserto control; L, ligasa.

**Tabla 2. Resumen de la caracterización de BACs con insertos experimentales.**

No. de transformantes	192 ufc/mL
Proporción de clonas con inserto	7/20 (35%)
Tamaño promedio del inserto	35.97±9.99Kb

**Tabla 3. Resultados de la secuenciación de los extremos del vector y análisis de las secuencias.**

Clonas secuenciadas	96
Pares de BES de calidad	59 (61.5%)
Total de bases secuenciadas	91,670
Promedio tamaño lectura (pb)	776.86±95.04

**Tabla 4. Tipos de pares de BES según sus parámetros de alineamiento.**

<b>Tipos de pares de BES</b>	<b>Cantidad</b>
De calidad	59
Con más de un alineamiento	0
Con un alineamiento	42
Concordantes	37
Verdaderos	21
Intermedios	2
Falsos	14
Opuestos	1
Indeterminados	4
Sin alineamiento	17

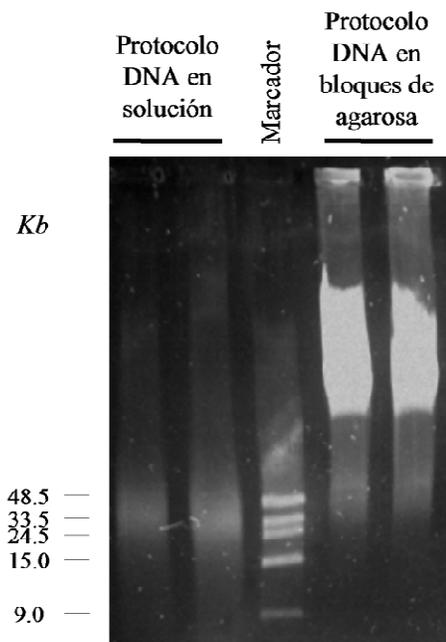
**Tabla 5. Resultados de los parámetros de alineamiento de los pares de BES concordantes, opuesto e indeterminados.**

<b>Tipo de pares de BES</b>		<b>Longitud promedio de alineamiento (pb)</b>	<b>Bit score promedio</b>	<b>Tamaño de inserto bioinformático promedio (pb)</b>
Concordantes	Verdaderos	732.90±127.21 n = 42	1408.33±269.46 n = 42	38,460.14±16,760.47 n = 21
	Intermedios	748.75±89.49 n = 4	1411.50±161.33 n = 4	7504 y 7621 n = 2
	Falsos	121.36±73.98 n = 28	236.68±146.58 n = 28	137.29±73.75 n = 14
Opuesto		788 y 804 n = 2	1557 n = 2	N.D.
Indeterminados		675.5±98.09 n = 4	1250.25±384.05 n = 4	N.D.

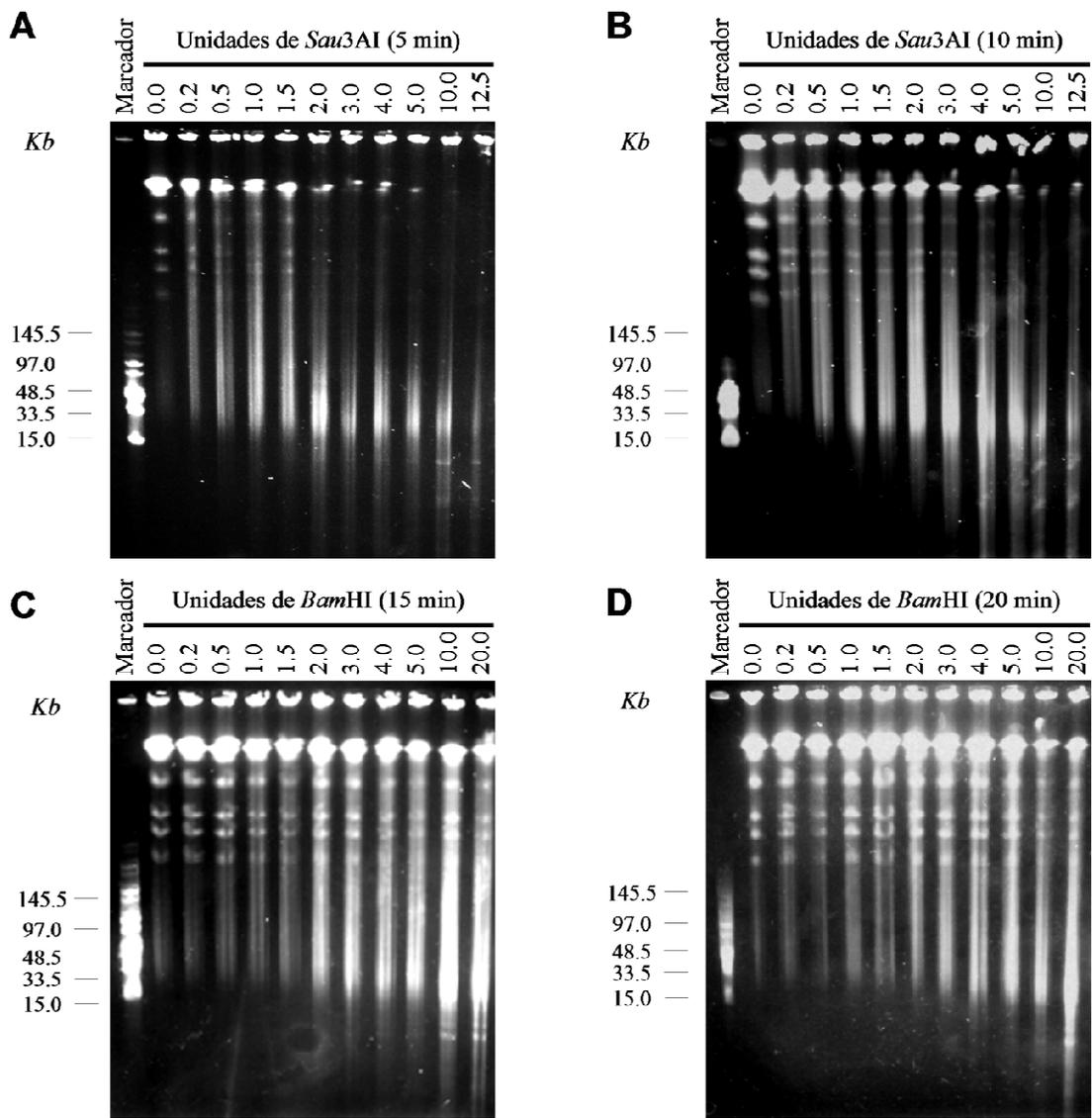
N.D.: no determinado.

**Tabla 6. Comparación de los tamaños de inserto bioinformáticos contra el tamaño de inserto estimado verificado experimentalmente de los distintos tipos de pares de BES.**

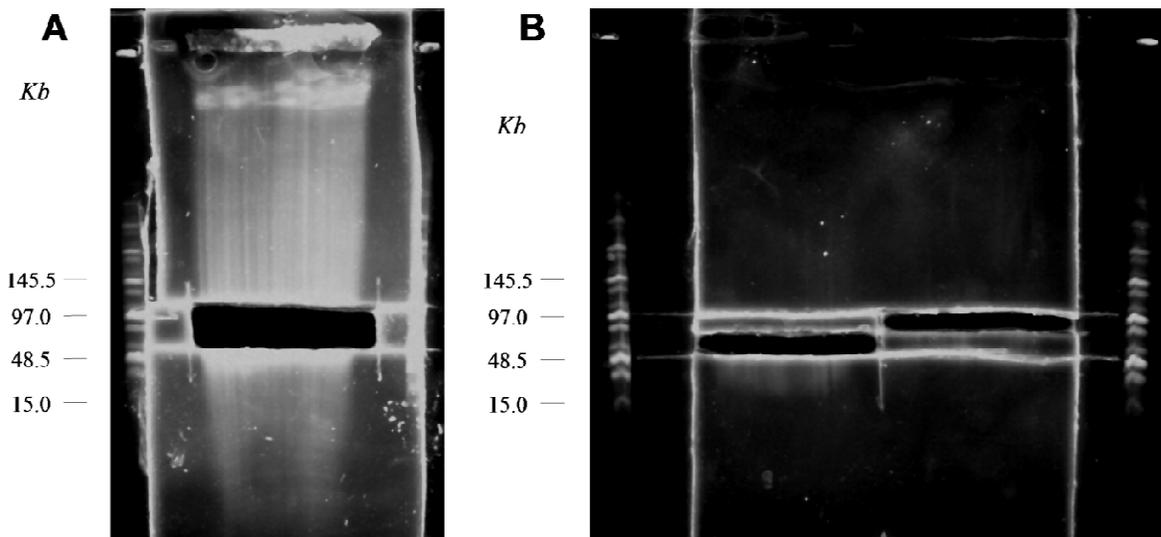
Tipo de pares de BES			Clona	Tamaño de inserto bioinformático (pb)	Tamaño de inserto estimado (pb)
Con un alineamiento	Concordantes	Verdaderos	A03	38,402	33,443
			A06	25,341	24,033
			C03	60,803	53,113
			D01	75,642	64,759
			G01	50,977	40,776
		Intermedios	A05	7,621	8,919
			G12	7,504	8,919
		Falsos	G08	172	Sin inserto
			H03	100	Sin inserto
	D12		74	Sin inserto	
	Indeterminados	B04	N.D.	28,958	
		F08	N.D.	20,988	
		G04	N.D.	36,444	
		H07	N.D.	34,102	
	Opuesto	F06	N.D.	25,674	
Sin alineamiento			A07	-	Sin inserto
			C09	-	Sin inserto
			G05	-	Sin inserto



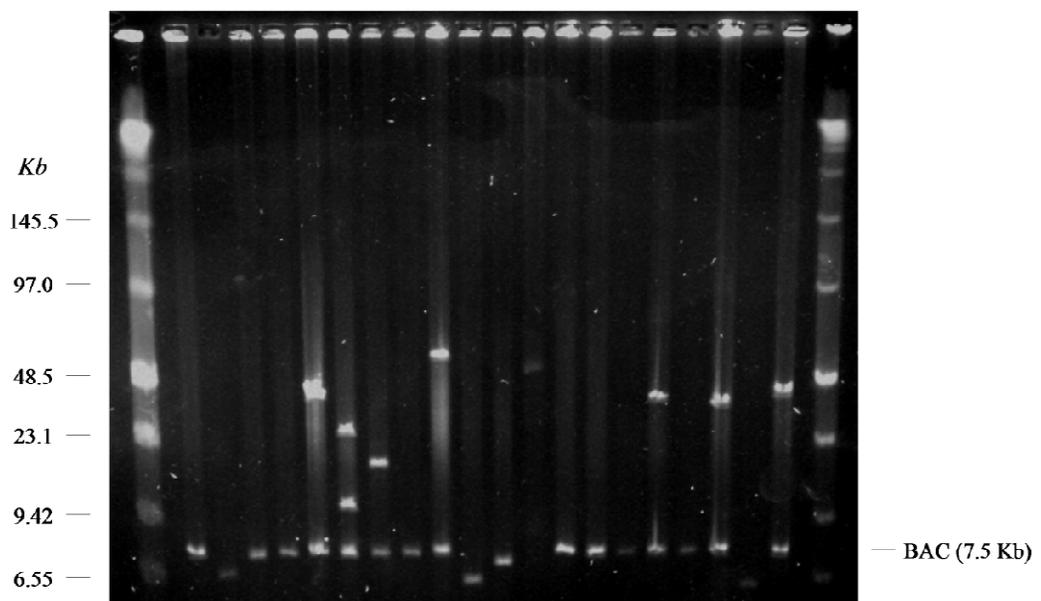
**Figura 1. Evaluación de la calidad del DNA obtenido usando dos protocolos de aislamiento.** El DNA aislado con el protocolo en bloques de agarosa fue de mayor peso molecular que el obtenido mediante el protocolo de DNA en solución. Marcador: DNA lambda ( $\lambda$ ,  $\lambda$ -*Xho*I,  $\lambda$ -*Xba*I).



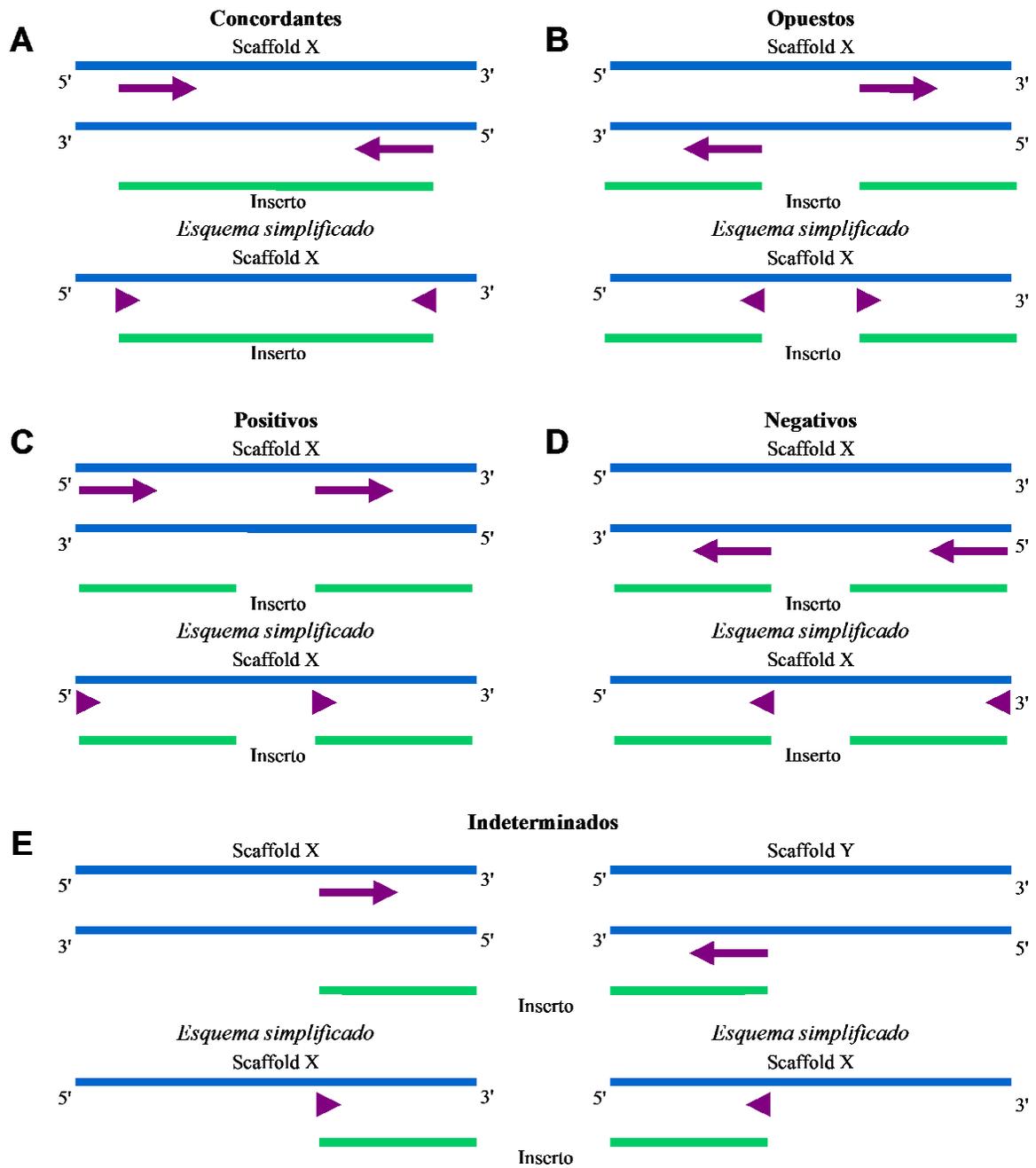
**Figura 2. Digestiones parciales de prueba del DNA experimental.** Digestiones con: (A) *Sau3AI* 5 min, (B) *Sau3AI* 10 min, (C) *BamHI* 15 min y (D) *BamHI* 20 min de incubación. En todos los casos el tamaño del DNA disminuyó conforme se aumentó la concentración de la enzima. Marcador: MidRange I PFG Marker de New England Biolabs.



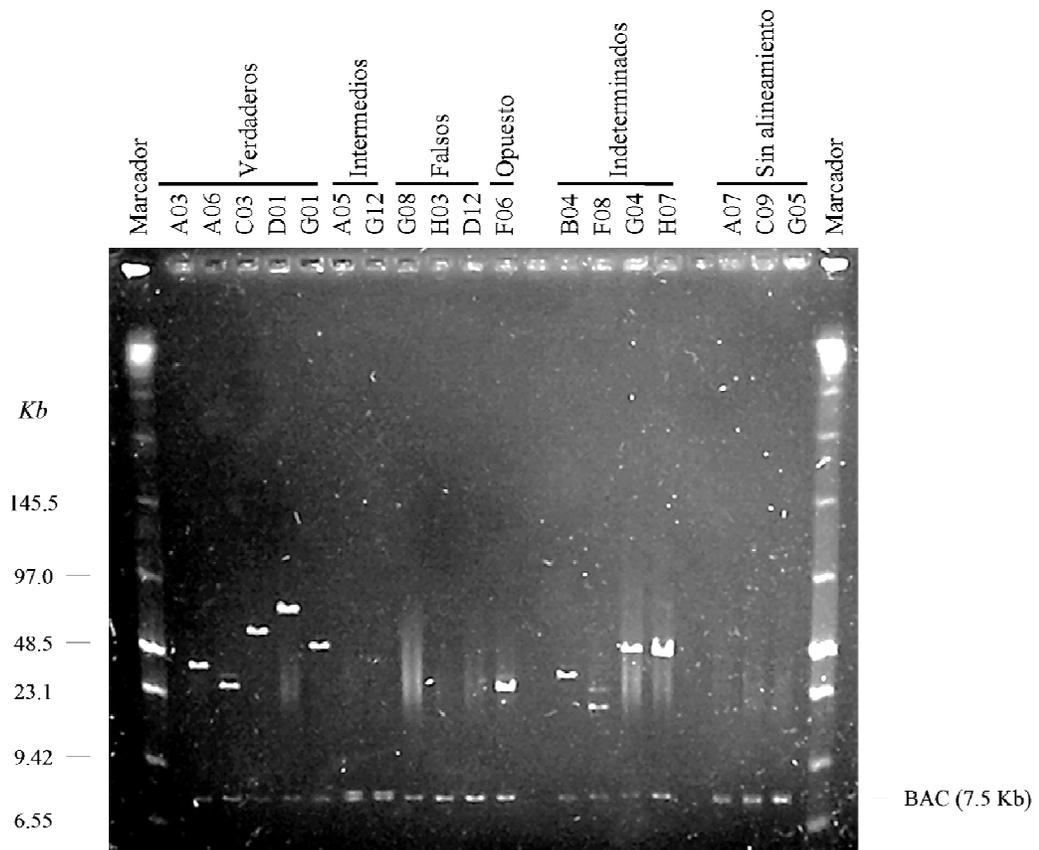
**Figura 3. Digestión en masa y doble selección de tamaño de insertos experimentales.** El DNA apareció como un barrido en la primera selección de tamaño (A), mientras que en la segunda selección (B) solo observamos algunos fragmentos por debajo de la región elegida. La porción en negro corresponde al fragmento del gel que se seleccionó antes de teñir el gel con bromuro de etidio para evitar daño al DNA. Marcador: MidRange I PFG Marker de New England Biolabs.



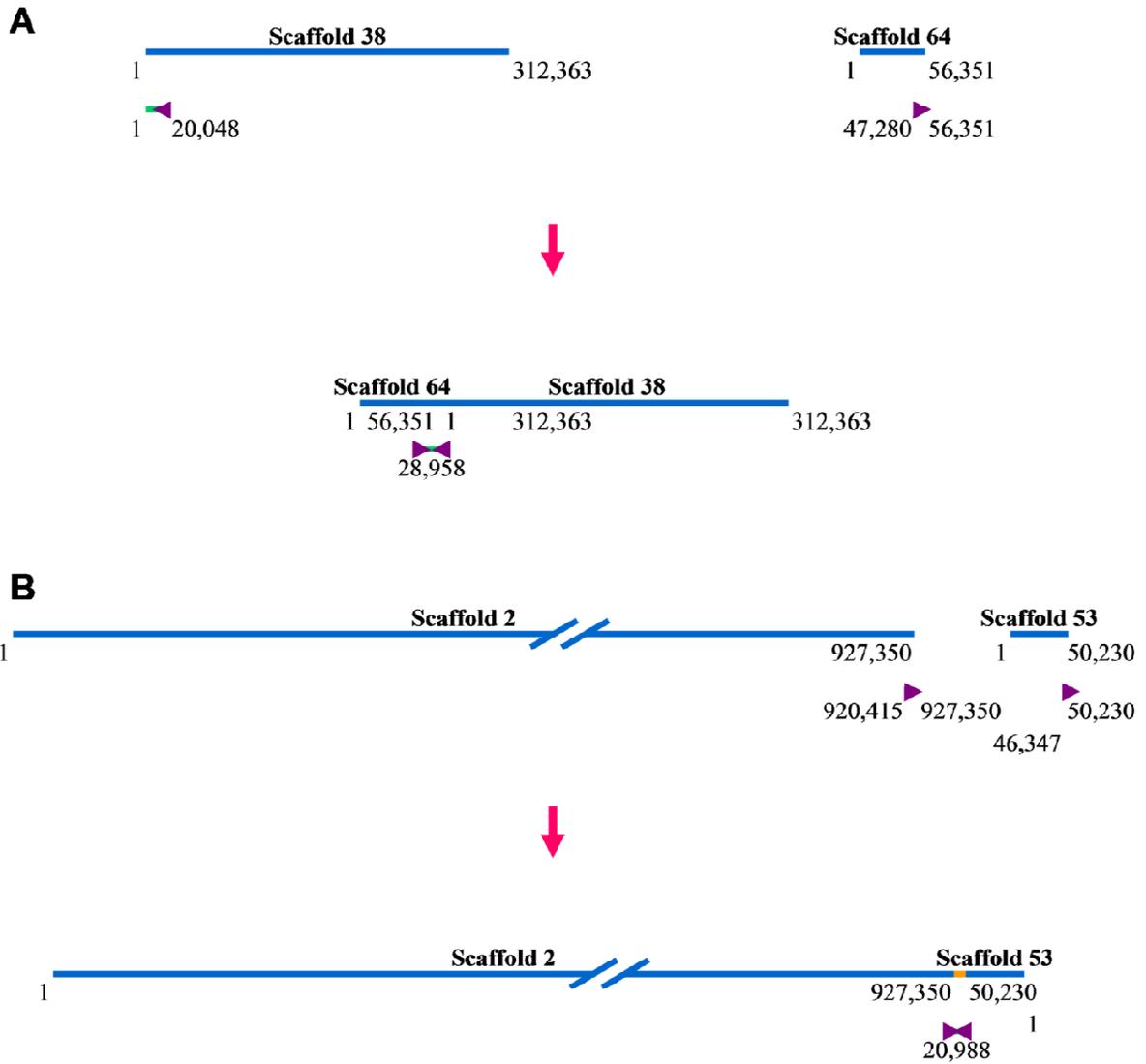
**Figura 4. Caracterización con *NotI* de clonas aleatorias de la transformación experimental.** Siete de las 20 clonas analizadas presentaron un inserto (35%). El tamaño promedio fue  $35.97 \pm 9.99$  Kb. Marcador: LowRange PFG Marker de New England Biolabs.



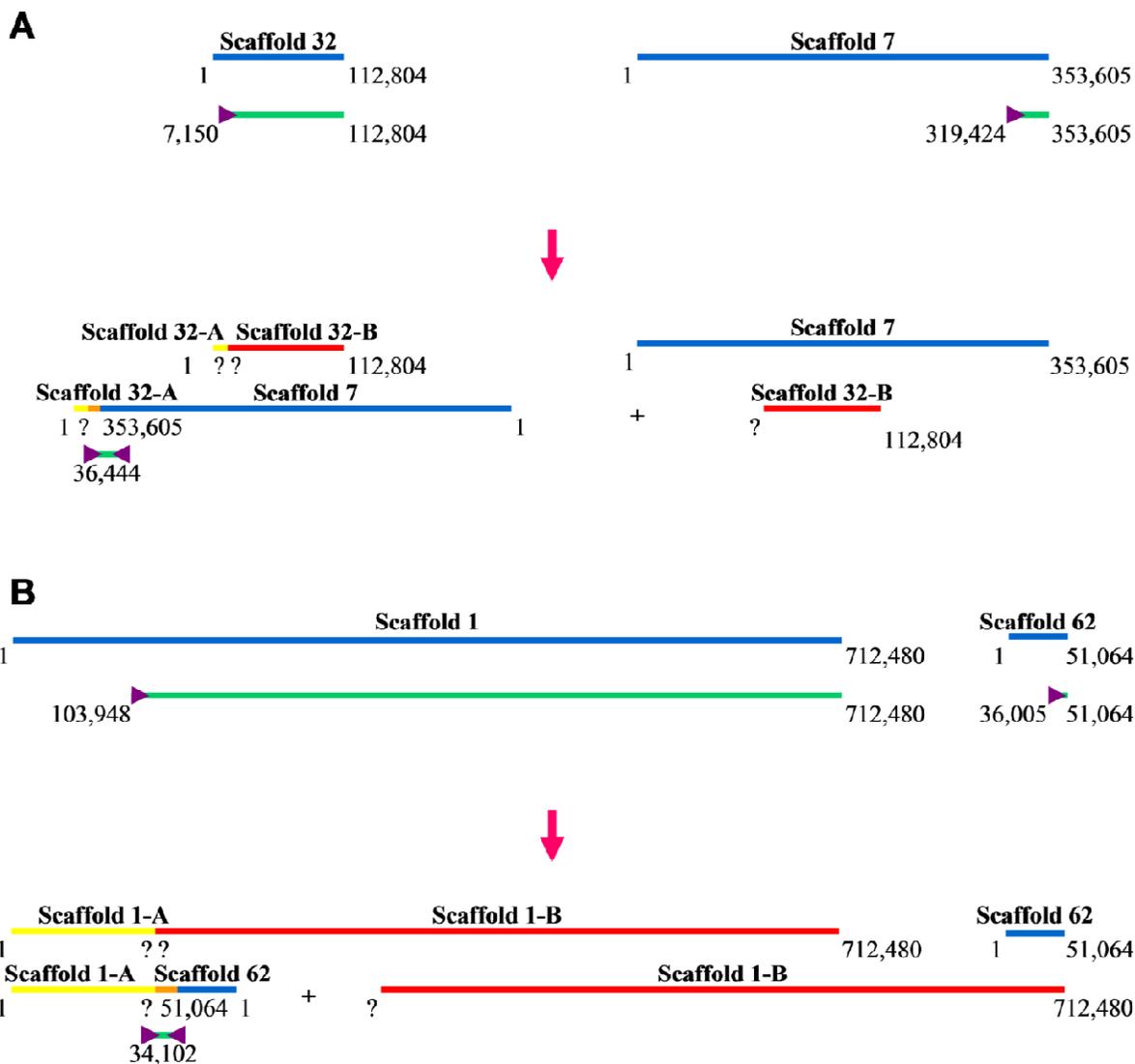
**Figura 5. Esquematisaciones de los distintos posibles alineamientos de los pares de BES.** Los alineamientos pueden ser concordantes (A) si son convergentes, opuestos (B) si son divergentes, positivos si tienen la misma dirección  $5' \rightarrow 3'$  (C), negativos si tienen la misma dirección  $3' \rightarrow 5'$  (D) e indeterminados si su dirección no se puede conocer debido a que alinean en scaffolds diferentes. Azul: scaffold; flechas moradas: BES; verde: inserto.



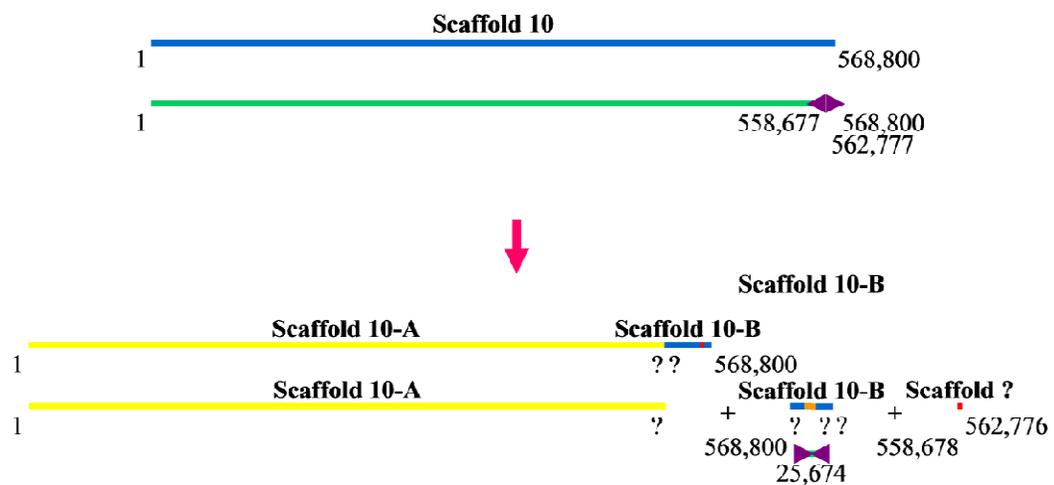
**Figura 6. Validaciones de los alineamientos y estimación del tamaño de inserto de los distintos tipos de pares de BES.** Las caracterizaciones de las clonas de pares de BES verdaderos, intermedios, falsos y sin alineamiento coincidieron con el resultado predicho por el alineamiento. El par de BES opuesto y los pares indeterminados arrojaron tamaños de inserto estimado que no se podían determinar con el alineamiento. Marcador: LowRange PFG Marker de New England Biolabs.



**Figura 7. Clonas que unen scaffolds diferentes.** Las clonas B04 (A) y F08 (B) son dos casos de pares de BES que alinearon en scaffolds diferentes y probablemente los unan. Azul: scaffold; flechas moradas: BES; verde: inserto; naranja: espacio.



**Figura 8. Clonas que detectan ensamblajes incorrectos.** Las clonas G04 (A) y H07 (B) son dos casos de pares de BES que alinearon en scaffolds diferentes pero con un tamaño de inserto mucho más grande al real estimado por lo que en lugar de unir scaffolds probablemente los reorganizaron y separaron. Azul: scaffold; flechas moradas: BES; verde: inserto; naranja: espacio; amarillo: porción probablemente conservada del scaffold; rojo: porción que probablemente no forman parte del scaffold.



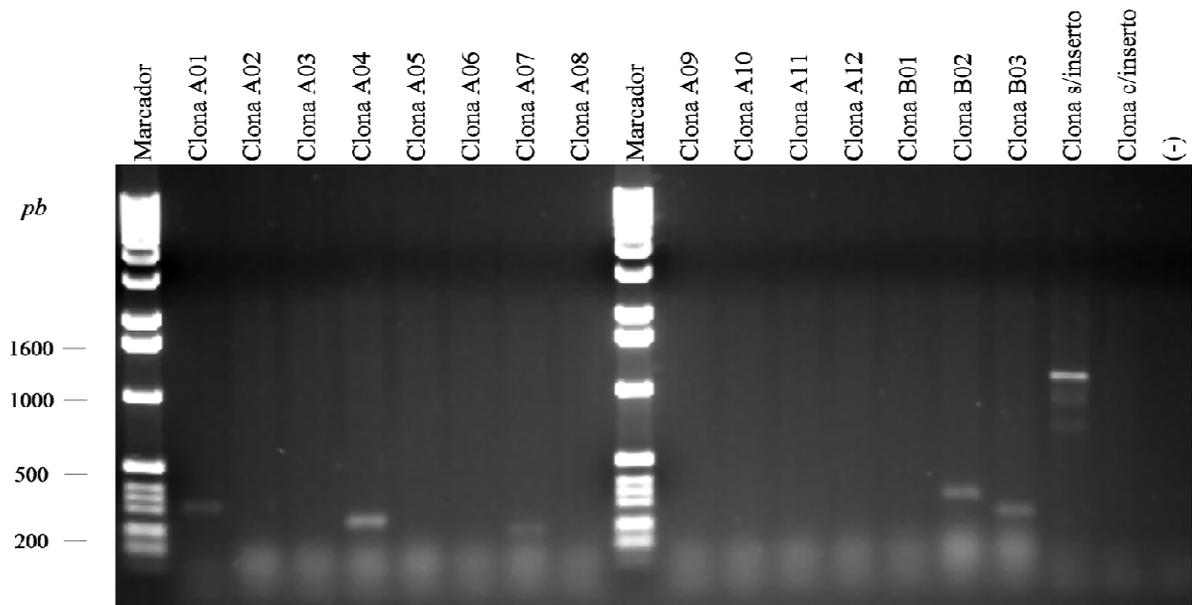
**Figura 9. Clona de par de BES opuesto.** La clona F06 tuvo pares de BES que alinearon en el mismo scaffold pero de manera opuesta lo cual sugiere que el scaffold 10 está unido por una región que no debería.

## **Anexos**

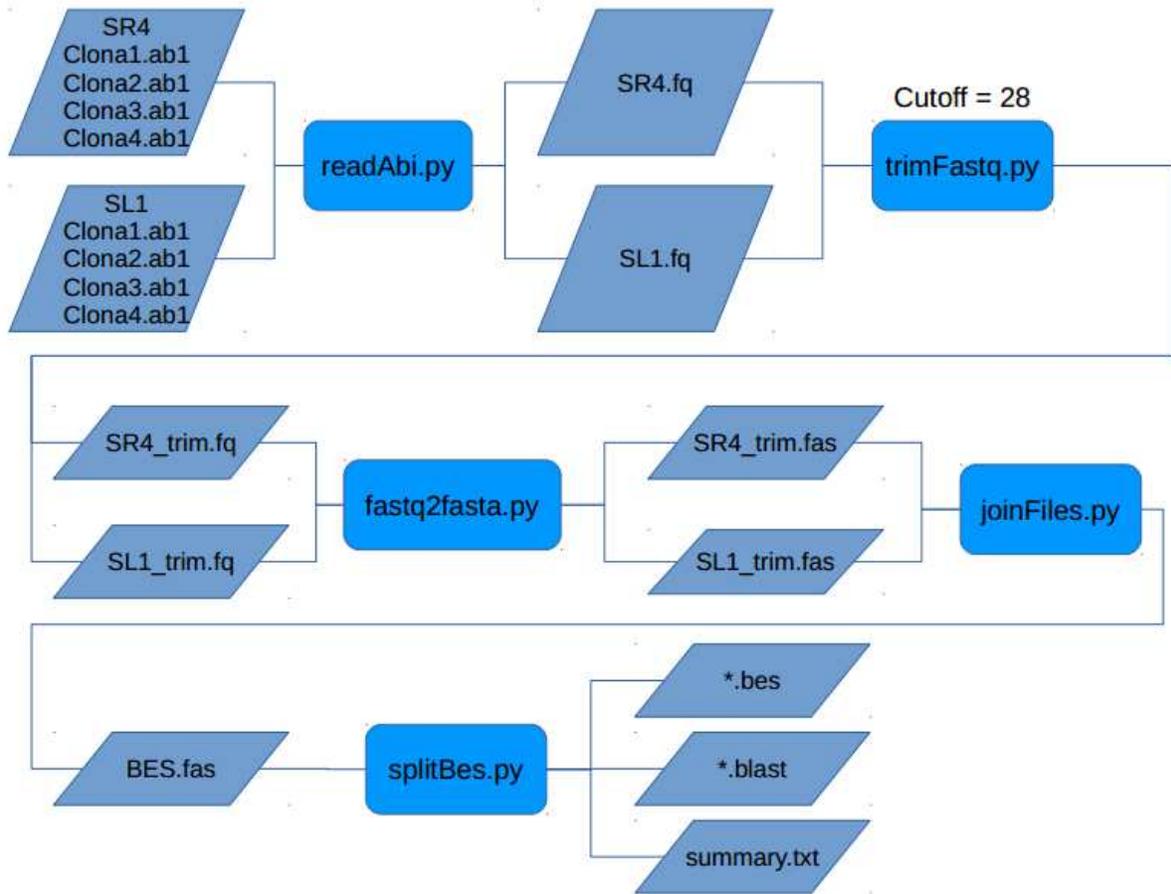
**Anexo 1. Protocolo de electroforesis de inversión de campo para la separación de fragmentos entre 3 a 70 kb en el equipo Pippin Pulse de Sage Science.**

Voltaje	75 V
A: Tiempo hacia adelante al inicio de la corrida	150 mseg
B: Tiempo hacia atrás al inicio de la corrida	50 mseg
C: Incremento añadido a A en cada paso	30 mseg
D: Incremento añadido a B en cada paso	10 mseg
E: Incremento añadido a C en cada paso	3 mseg
F: Incremento añadido a D en cada paso	1 mseg
G: Número de pasos por ciclo	48
Tiempo de corrida	14 h

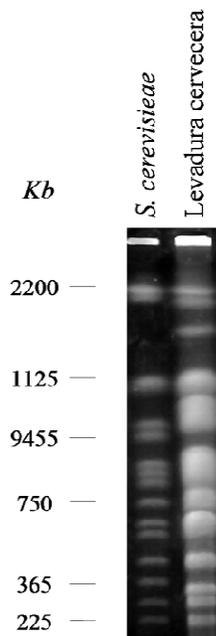




**Anexo 3. Detección de clonas sin inserto mediante PCR de colonia.** La presencia de una banda indica una clona sin inserto. Marcador: 1 Kb Plus DNA Ladder de Invitrogen; (-): control negativo reacción.



**Anexo 4. Diagrama de flujo del análisis de las secuencias de los extremos del vector.**



**Anexo 5. Electrocariotipo de *Saccharomyces cerevisiae* y la levadura cervecera.** PFGE en un gel de agarosa para campos pulsados de Molecular Sigma Biology (agarosa-TBE 0.5X al 1 %) con un protocolo para separar cromosomas de levadura en el equipo CHEF Mapper XA de Bio-Rad (6.0 V/cm, 120°, 60-120 s, lineal, 24 h, 14°C). El cromosoma más pequeño de la levadura cervecera es de alrededor de 200 kb.

**Anexo 6. Lista de los pares de BES con alineamiento contra los scaffolds del ensamblaje de la levadura cervecera y el cromosoma de *Saccharomyces cerevisiae* o *Saccharomyces bayanus* contra el que alinean dichos scaffolds.** Cromosomas de *Saccharomyces cerevisiae*: Sc; cromosomas de *Saccharomyces bayanus*: Sb.

BES	Scaffold	Tipo de BES	Inserto bioinformático	Cromosoma
Placa_1A03-SR4	scaffold22 size618977	Verdadero	38402	Sc2
Placa_1A03-SL1	scaffold22 size618977	Verdadero	38402	Sc2
Placa_1A06-SR4	scaffold38 size312363	Verdadero	25341	Sc16
Placa_1A06-SL1	scaffold38 size312363	Verdadero	25341	Sc16
Placa_1A10-SR4	scaffold13 size669881	Verdadero	39553	Sc11
Placa_1A10-SL1	scaffold13 size669881	Verdadero	39553	Sc11
Placa_1A11-SR4	scaffold59 size153105	Verdadero	32685	Sc8
Placa_1A11-SL1	scaffold59 size153105	Verdadero	32685	Sc8
Placa_1B01-SR4	scaffold15 size969117	Verdadero	27726	Sb13
Placa_1B01-SL1	scaffold15 size969117	Verdadero	27726	Sb13
Placa_1B07-SR4	scaffold9 size551088	Verdadero	37260	Sc14
Placa_1B07-SL1	scaffold9 size551088	Verdadero	37260	Sc14
Placa_1B08-SR4	scaffold14 size919676	Verdadero	11210	Sb2 y Sb4
Placa_1B08-SL1	scaffold14 size919676	Verdadero	11210	Sb2 y Sb4
Placa_1C03-SR4	scaffold3 size1404408	Verdadero	60803	Sc13 y Sb15
Placa_1C03-SL1	scaffold3 size1404408	Verdadero	60803	Sc13 y Sb15
Placa_1D01-SR4	scaffold26 size991323	Verdadero	75642	Sb2 y Sb4
Placa_1D01-SL1	scaffold26 size991323	Verdadero	75642	Sb2 y Sb4
Placa_1D03-SR4	scaffold38 size312363	Verdadero	25348	Sc16
Placa_1D03-SL1	scaffold38 size312363	Verdadero	25348	Sc16
Placa_1E05-SR4	scaffold22 size618977	Verdadero	45748	Sc2
Placa_1E05-SL1	scaffold22 size618977	Verdadero	45748	Sc2
Placa_1E08-SR4	scaffold7 size353605	Verdadero	28239	Sc8
Placa_1E08-SL1	scaffold7 size353605	Verdadero	28239	Sc8
Placa_1E11-SR4	scaffold29 size967628	Verdadero	26965	Sc4
Placa_1E11-SL1	scaffold29 size967628	Verdadero	26965	Sc4
Placa_1F07-SR4	scaffold3 size1404408	Verdadero	71726	Sc13, Sb8 y Sb15
Placa_1F07-SL1	scaffold3 size1404408	Verdadero	71726	Sc13, Sb8 y Sb15
Placa_1F09-SR4	scaffold15 size969117	Verdadero	49112	Sb13
Placa_1F09-SL1	scaffold15 size969117	Verdadero	49112	Sb13
Placa_1G01-SR4	scaffold26 size991323	Verdadero	50977	Sb2 y Sb4
Placa_1G01-SL1	scaffold26 size991323	Verdadero	50977	Sb2 y Sb4
Placa_1G02-SR4	scaffold22 size618977	Verdadero	37744	Sc2
Placa_1G02-SL1	scaffold22 size618977	Verdadero	37744	Sc2
Placa_1G03-SR4	scaffold16 size319100	Verdadero	12593	Sb2 y Sb4
Placa_1G03-SL1	scaffold16 size319100	Verdadero	12593	Sb2 y Sb4
Placa_1G06-SR4	scaffold25 size527137	Verdadero	45190	Sb5
Placa_1G06-SL1	scaffold25 size527137	Verdadero	45190	Sb5
Placa_1G11-SR4	scaffold3 size1404408	Verdadero	27650	Sc13, Sb8 y Sb15

Placa_1G11-SL1	scaffold3 size1404408	Verdadero	27650	Sc13, Sb8 y Sb15
Placa_1H01-SR4	scaffold22 size618977	Verdadero	37749	Sc2
Placa_1H01-SL1	scaffold22 size618977	Verdadero	37749	Sc2
Placa_1A05-SR4	scaffold13 size669881	Intermedio	7621	Sc11
Placa_1A05-SL1	scaffold13 size669881	Intermedio	7621	Sc11
Placa_1G12-SR4	scaffold8 size435656	Intermedio	7504	Sc9
Placa_1G12-SL1	scaffold8 size435656	Intermedio	7504	Sc9
Placa_1F06-SR4	scaffold10 size568800	Opuesto	N.D.	Sc12 y Sb12
Placa_1F06-SL1	scaffold10 size568800	Opuesto	N.D.	Sc12 y Sb12
Placa_1B04-SR4	scaffold64 size56351	Indeterminado	N.D.	Sc16
Placa_1B04-SL1	scaffold38 size312363	Indeterminado	N.D.	Sc16
Placa_1F08-SR4	scaffold53 size50230	Indeterminado	N.D.	Sc15
Placa_1F08-SL1	scaffold2 size927350	Indeterminado	N.D.	Sc15
Placa_1G04-SR4	scaffold7 size353605	Indeterminado	N.D.	Sc8
Placa_1G04-SL1	scaffold32 size112804	Indeterminado	N.D.	Sc3 y Sc5
Placa_1H07-SR4	scaffold1 size712480	Indeterminado	N.D.	Sb1 y Sb12
Placa_1H07-SL1	scaffold62 size51064	Indeterminado	N.D.	Sb1

**Anexo 7. Resumen rearrreglos propuestos y cromosomas de *Saccharomyces cerevisiae* o *Saccharomyces bayanus* contra el que alinean los scaffolds individuales de cada rearrreglo.**  
 Cromosomas de *Saccharomyces cerevisiae*: Sc; cromosomas de *Saccharomyces bayanus*: Sb.

<b>BES</b>	<b>Scaffold</b>	<b>Rearreglo propuesto</b>	<b>Cromosoma</b>
Placa_1B04-SR4	scaffold64	Scaffold 64-38	Sc16
Placa_1B04-SL1	scaffold38		Sc16
Placa_1F08-SR4	scaffold53	Scaffold 2-53	Sc15
Placa_1F08-SL1	scaffold2		Sc15
Placa_1G04-SR4	scaffold7	Scaffold 32A-7 + Scaffold 32B	Sc8
Placa_1G04-SL1	scaffold32		Sc3 y Sc5
Placa_1H07-SR4	scaffold1	Scaffold 1A-62 + Scaffold 1B	Sb1 y Sb12
Placa_1H07-SL1	scaffold62		Sb1