

# Constructing large DNA segments by iterative clone recombination

Duane E. Smailus · Rene L. Warren ·  
Robert A. Holt

Received: 25 October 2007 / Revised: 28 November 2007 / Accepted: 8 January 2008 / Published online: 24 January 2008  
© Springer Science+Business Media B.V. 2008

**Abstract** Methods for constructing large contiguous segments of DNA will be enabling for Synthetic Biology, where the assembly of genes encoding circuits, biosynthetic pathways or even whole microbial organisms is of interest. Currently, in vitro approaches to DNA synthesis are adequate for generating DNAs that are up to 10s of kbp in length, and in vivo recombination strategies are more suitable for building DNA constructs that are 100 kbp or larger. We have developed a vector system for efficient assembly of large DNA molecules by iterative in vivo recombination of fosmid clones. Two custom fosmid vectors have been built, pFOSAMP and pFOSKAN, that support antibiotic switching. Using this technique we rebuilt two non-contiguous regions of the *Haemophilus influenzae* genome as episomes in recombinogenic *Escherichia coli* host cells. These regions together comprise 190 kbp, or 10.4% of the *H. influenzae* genome.

**Keywords** DNA synthesis · Gene synthesis · Synthetic genome · Recombineering · Lambda Red · Fosmid

## Introduction

Current approaches to gene synthesis utilize short oligonucleotides that are subjected to iterative rounds of ligation and amplification to make longer DNA sequences. This method has been used successfully to build complete coding sequences for individual genes, gene clusters (Tian et al. 2004), and even complete viral genomes (Cello et al. 2002; Smith et al. 2003). The power of this approach, in terms of bases of DNA synthesized per unit cost appears to be advancing on a trajectory reminiscent of Moore's law (Carlson 2003). In vitro approaches for generating much larger sequences are compromised by the fragility high molecular weight DNA and the diminishing yield of the desired product as length increases. Thus, the construction of large DNA molecules has generally relied on recombination of precursor fragments in a host organism (Itaya et al. 2005). The lambda Red system (Yu et al. 2000) is an efficient and scarless method of in vivo recombination. This system utilizes a host strain (typically *E. coli* DH10B cells) carrying a segment of the phage lambda genome that contains the *exo*, *bet* and *gam* genes under control of a temperature-sensitive repressor. These lambda genes mediate recombination between the ends of a linear incoming DNA segment with homologous sequences in a target DNA. The homology regions can be very short (~50 bp) and the target can be any chromosomal or episomal DNA molecule present in the host cell. We have designed a fosmid vector system that allows lambda Red recombinations to be done iteratively, such that large DNAs can be assembled stepwise in the host cell. For proof of principle we are reassembling segments of the *H. influenzae* Rd KW20 genome in an *E. coli* host, using a recombination strategy similar in principle to that described by Kotzamanis and Huxley (2004). Briefly, a genomic library

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s11693-008-9011-6) contains supplementary material, which is available to authorized users.

---

D. E. Smailus · R. L. Warren · R. A. Holt (✉)  
Canada's Michael Smith Genome Sciences Centre, British  
Columbia Cancer Agency, Suite 100, 570 West 7th Avenue,  
Vancouver, BC, Canada V5Z 4S6  
e-mail: rholt@bcgsc.ca

is constructed in each of two fosmid vectors that carry different antibiotic resistance markers. Clones are randomly sequenced, mapped to the genome and a minimal tiling set of genomic fragments in alternating vectors is selected. DNA for a given clone is isolated, linearized, and transformed into cells containing the next clone in the tiling set. A lambda RED mediated fusion event joins the incoming fragment with its neighbour and replaces the antibiotic resistance gene, allowing selection of recombinants. The process is then repeated. Here we demonstrate the re-assembly by iterative recombination of two non-contiguous *H. influenzae* gDNA segments totaling of 190 kbp, or 10.4% of the genome. In principle, this approach could be used to rebuild and reboot the complete *H. influenzae* genome. This is our longer-term goal with this model system, but requires consideration of additional issues such as cross-talk in gene expression and incompatibility of some gene products (Holt et al. 2007). The method also general utility in constructing large DNAs.

## Material and methods

### Vector construction

Two fosmid vectors with different antibiotic resistance genes were constructed. Fosmids are single copy vectors that use the F replicon. pEpiFOS5 (Epicentre)(Genbank EU140753.1), a derivative of BAC vector pBeloBAC11, was digested with Eco72I and ScaI to remove a 500 bp segment between these sites that showed exact homology to the *E. coli* genome (and could, therefore, interfere with recombination). The 500 bp segment was replaced by a DNA segment comprised of an FseI restriction site (for which there are no sites in the *H. influenzae* genome), a SmaI restriction site (a rare 8 bp cutter for which there are no other sites in the vector backbone, and which provides blunt ends for cloning), and either the ampicillin resistance gene (from plasmid pET19b) or the kanamycin resistance gene (from TN7). SmaI is the insert site, and the FseI site is used to linearize the vector prior to recombination. The correct assembly of these two new vectors, pFOSAMP and pFOSKAN was verified by sequencing. Genbank identifiers are EU292739 and EU292740.

### Library construction

*H. influenzae* Rd KW20 cells were cultured overnight in brain-heart infusion broth supplemented with hemin (10 µg/ml) and nicotinamide adenine dinucleotide (100 µg/ml). Cells were pelleted and resuspended in Lysis Solution (10 mM Tris-Cl pH 8, 100 mM EDTA pH 8,

0.5% (w/v) SDS, 20 µg/ml RNase A, Proteinase K 100 µg/ml and incubated in a 50°C bath, 3 h, with mixing by gentle inversion every hour. Lysate was extracted three times with equal volumes of phenol:chloroform:isoamyl alcohol (25:24:1) then ethanol precipitated, spooled, and dissolved in TE. DNA was hydrodynamically sheared with a 25 gauge needle and a 25–40 kbp size fraction was isolated by pulsed field gel electrophoresis (PFGE). Using the EpiFos Fosmid Library Production Kit (Epicentre), size-selected *H. influenzae* DNA was cloned into SmaI-linearized and dephosphorylated pFOSAMP and pFOSKAN vectors, then packaged and plated on 2XYT agar containing the appropriate antibiotic, according to the manufacturer's instructions.

### Clone mapping

Fosmid end sequences were obtained using custom primers (pFOSKAN\_forward 5'>GAGCATTACGCTGACTTGAC; pFOSAMP\_forward 5'>ACGATAGTTACCGGATAAGG; reverse 5'>CAAATATTATACGCAAGGCG) and previously described nanolitre scale Sanger sequencing methods (Smailus et al. 2005). A total of 11,520 total fosmid paired end sequences (5,760 from each library) were obtained and these were vector-trimmed using cross\_match (<http://www.phrap.org>) and quality trimmed using trim2 (-M 10) (Huang et al. 2003). The resulting 9,935 sequences (5,034 from the pFOSAMP library and 4,901 from the pFOSKAN library) were aligned to the *H. influenzae* Rd KW20 reference genome sequence (NC\_000907) using wuBLASTn (blast version 2.0, May 10th, 2005; <http://www.blast.wustl.edu>). The default parameters were used and only the best scoring match from each fosmid read with alignments longer than 200 nucleotides and sharing more than 70% sequence identity with the reference genome were subsequently evaluated. While mapping fosmid end-reads, we ensured that the pairing logic was respected, with pairs from any given clones aligning in opposite directions, facing inwards. Pairs aligning outside 40 kbp ± 2 SD of the insert size distribution were not considered. Custom software was designed to aid in mapping the genomic constructs onto the complete *H. influenzae* genome sequence (NC\_000907) and to help identify suitable candidates for the minimal tiling set.

### Selecting minimal tiling set

Because our iterative recombination scheme is directional, all tiling path clones must have inserts that map to the same strand (e.g. they must be in the same orientation with respect to vector). Higher coverage (99.21%) was obtained for the plus strand than the minus strand, so clones for the

minimal tiling set were selected from the plus strand. We established the following rules for selecting a minimum set of clones; (1) overlapping clones must have alternate selectable markers, (2) the 3'-most 50–100 bp of the linearized incoming clone must not align to the *E. coli* genome, or to any repeats within the *H. influenzae* genome, and (3) clone inserts should provide maximal genome coverage and show minimal overlap, with a suitable overlap ~500–10,000 bp, but no smaller than 50 bp. Intra and inter chromosomal repeats were detected using `cross_match` (<http://www.phrap.org>) and repeats larger than 20 bp and having more than 70% sequence identity were avoided. Selection was performed in a semi-automated fashion, whereby suitable clones were flagged by software written in-house and validated manually, putting emphasis on the uniqueness of 3' end sequences while ensuring a maximal clone overlap for mediating recombination. The final minimal tiling set included 61 clones, 31 from the pFOS-AMP library and 30 from the pFOSKAN library. The tiling set includes three gaps (genome coordinates 145,915–157,386; 1,275,539–1,290,904; and 1,508,825–1,510,549 bp) and covers 98.5% of the 1.83 Mbp *H. influenzae* genome. Cloning ribosomal RNA genes has been problematic in other systems (Itaya et al. 2005), but the unclonable regions we encountered did not contain or intersect with any *H. influenzae* ribosomal RNA genes.

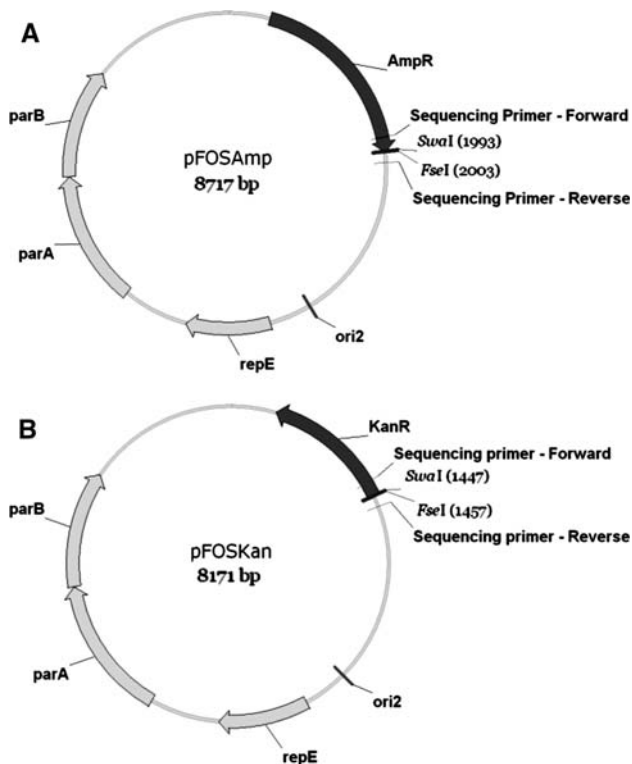
#### Iterative clone assembly

We report iterative assembly of two regions of the genome in this proof of principle study. For each region, the first of the three fosmids to be assembled was transformed into EL350 cells (Lee et al. 2001) which harbour the prophage encoding lambda recombination proteins *exo*, *bet* and *gam* under control of the cI857 temperature sensitive transcriptional repressor. These cells were cultured, heat-shocked for 15 min at 42°C, then immediately cooled on ice and made electrocompetent by washing three times with ice-cold 10% glycerol. The second clone in each set of three was linearized by FseI digestion and purified by PFGE to remove any uncut DNA. Purified DNA was then end-treated to remove the residual bases of the FseI site using Bal31 exonuclease, and then end-polished using T4 polymerase/T4 Kinase. Bal31 digestion was done for 2 min at room temperature in a total reaction volume of 200 µl, that contained approximately 2 µg DNA, 1 unit of Bal31 (New England Biolabs) and 1× Bal31 buffer (New England Biolabs). DNA was purified by phenol:chloroform:isoamyl alcohol (25:24:1) extraction and ethanol precipitation, resuspended in 10 µl of Tris EDTA (10 mM Tris, 1 mM EDTA, pH 8) then end-polished with T4 polymerase (Epicentre) and T4 polynucleotide kinase (Epicentre) to

give 5'-phosphorylated blunt ends. The end polishing reaction contained approximately 2 µg DNA, 1 mM dNTPs, 1 mM ATP, 1× end repair buffer (Epicentre) and 1 µl of T4 polymerase/T4 kinase enzyme mix (Epicentre). The reaction was incubated at room temperature for 30 min then terminated by incubation at 70°C for 10 min. End polished DNA was recovered by phenol:chloroform:isoamyl alcohol (25:24:1) extraction and ethanol precipitation, resuspended in 5 µl of Tris EDTA (10 mM Tris, 1 mM EDTA, pH 8). Two microliters of linearized end-polished fosmid DNA was combined with 50 µl of heat shocked electrocompetent EL350 cells and transferred to a pre-chilled 1 mm electroporation cuvette (Biorad). Electroporation was performed using a Biorad GenePulser, 50 µF, 1.8 kV. Immediately after electroporation cells were suspended in 450 µl SOC medium and incubated at 32°C for 1 h. Recombinants were selected by plating the entire transformation on 2XYT solid media containing antibiotic matching the resistance marker on the incoming linearized clone and incubating overnight at 32°C. For these experiments we typically observe up to 12 colonies per recombination. All colonies are screened by end-sequencing and restriction digestion and we typically observe one or two successful recombinants per attempt.

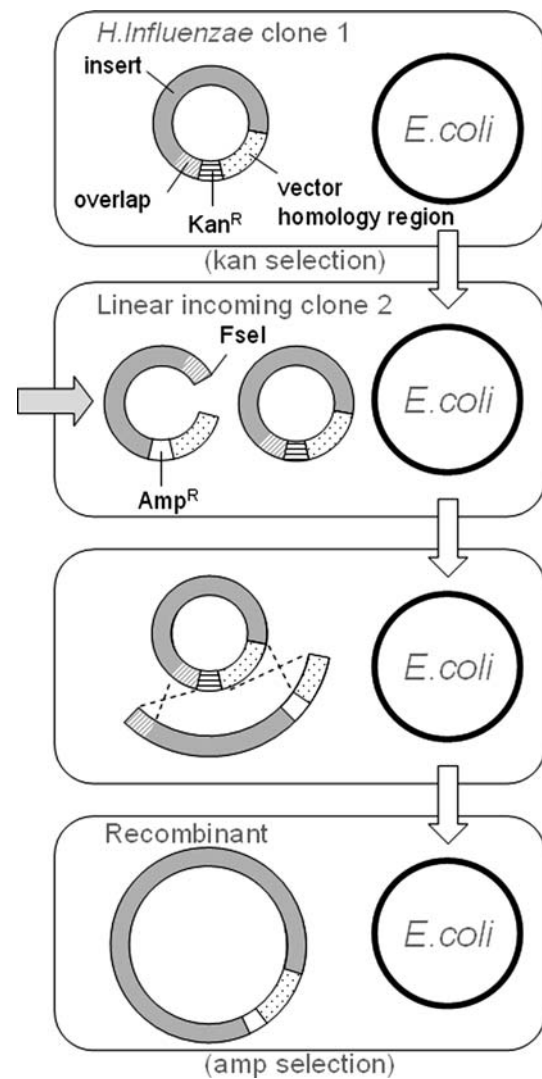
#### Results

Two modified fosmid vectors—one containing an ampicillin resistance marker (pFOSAMP, Fig. 1a) and the second a kanamycin resistance marker (pFOSKAN, Fig. 1b) were constructed by modifying the pEPIFOS5 fosmid vector backbone. A unique FseI restriction site was engineered to the right of the insert site and serves to linearize the clones that are to recombine with the target construct in the host. Two *H. influenzae* fosmid libraries were built, one in each vector, and random clones were end-sequenced to high redundancy. End-sequences were mapped to the *H. influenzae* Rd KW20 reference genome sequence and repeats within and between *H. influenzae* and *E. coli* were highlighted to facilitate selection of a minimal tiling set of clones with non-repetitive end sequences. The selected minimal tiling set included 61 clones, 31 from the pFOSAMP library and 30 from the pFOSKAN library and covers 98.5% of the 1.83 Mbp *H. influenzae* genome. Three segments of the genome (coordinates 145,915–157,386; 1,275,539–1,290,904; and 1,508,825–1,510,549 bp) were not spanned by any clone and, therefore, these relatively small physical gaps are not represented in the tiling set. We demonstrate successful iterative recombination by reassembling independently two non-contiguous regions of the *H. influenzae* genome using the method illustrated in Fig. 2. For each region, the first of three overlapping



**Fig. 1** Maps of custom fosmid cloning vectors pFOSAMP and pFOSKAN. Each vector contains a unique blunt-end cloning site (*Swa*I) and unique restriction site (*Fse*I) for linearization of clones prior to recombination

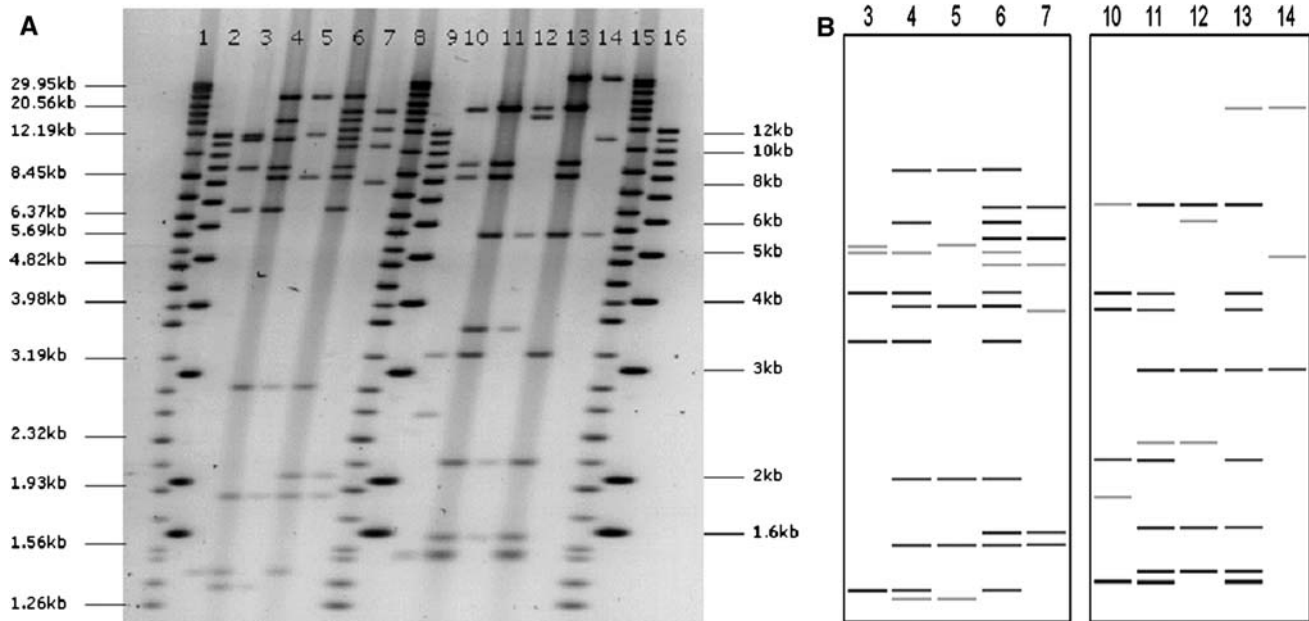
fosmids was transformed into lambda Red compatible *E. coli* cells, cultured in the presence appropriate antibiotic (e.g. ampicillin) and electrocompetent cells were prepared. The next clone in the tiling path was linearized by *Fse*I, end polished, then transformed into the cells containing the initial fosmid. The incoming clone was joined with the first clone by a recombination event mediated at one end by the vector sequence and at the other end by overlapping genomic sequence. A key feature of this system is that upon recombination, the new genomic segment is joined to the first, but at the same time the vector segment from the first clone is replaced with the vector of the second clone. As such, the product contains only a single vector sequence, but now the antibiotic resistance marker has been exchanged and recombinant clones can be selected. Importantly, the incoming clone cannot propagate on its own because it is linear, and only becomes circularized (and thus replicable) upon recombination with the first resident clone. A second round of recombination was then undertaken with the third fosmid clone for each region, to generate large contiguous *H. influenzae* genomic DNA segments propagating in their *E. coli* hosts. The size and content of the intermediate and final constructs was verified by *Eco*RI restriction mapping (Fig. 3) and also by sizing the *Fse*I cut gDNA by PFGE (Fig. S1). End sequences of



**Fig. 2** Fusion of fosmid clones by lambda Red recombination. The ends of an incoming linear clone (ampicillin resistant) recombine with homologous sequence in a resident circular clone (kanamycin resistant). One region of end homology is between vectors (dotted) and the other between insert ends (light grey). Recombinants are selected according to the marker on the incoming clone (kanamycin)

each final construct verify the reassembly of test segment 1 (bases 446,461–553,659) and test segment 2 (bases 63,192–145,915 bp) of the *H. influenzae* chromosome. Regarding the success rate of these types of recombination procedures it is important to note that we began initially with 11 sets of three clones that in total spanned half of the *H. influenzae* genome. Successful pairwise recombinations were obtained for five of these sets. The two sets reported above were the first to undergo successfully the second round of recombination, and thereby demonstrate the feasibility of iterative recombination. Recombination failures could be related to scale, where an inadequate number of cells were plated and screened, or perhaps to reconstitution





**Fig. 3** EcoRI restriction maps of individual fosmid clones that are combined to form larger contiguous DNA inserts. There are three fosmid clones covering each of two regions of the *H. influenzae* genome. For each set of three fosmids, two iterative recombination events are required for assembly. Panel **a** shows EcoRI digests of the initial fosmids (lanes 3, 5 and 7 for region 1, and lanes 10, 12 and 14 for region 2) used in reconstruction. Digests of intermediate clones created by fusing the first two clones in each set are shown in lane 4 (fusion of clones from lanes 3 and 5) and lane 11 (fusion of clones from lanes 10 and 12). Digests of the final constructs created by fusing each intermediate clone with the third and final clone in each set are shown in lane 6 (fusion of clones from lanes 5 and 7) and lane

13 (fusion of clones from lanes 12 and 14). Lanes 1, 2, 8, 9, 15 and 16 contain size markers. Panel **b** shows the expected banding patterns from in-silico digestion of the constructs from Panel **a**. The lanes in Panel **b** match those in Panel **a**. Grey bands in Panel **b** are restriction fragments that contain vector DNA. Each DNA sample was prepared by alkaline lysis from a 1.2 ml overnight culture, then digested overnight with 10 U of EcoRI (New England Biolabs) in a 10  $\mu$ l reaction volume. The digests were run on a SeaKem LE agarose (Cambrex) gel in 1 $\times$  TAE buffer, 200 V, for 5 h. The SYBR green stained gel was scanned using a FluorImager 595 (Molecular Dynamics)

of toxic gene combinations incompatible with host cell viability, as discussed below.

## Discussion

The ability to construct large DNA's that represent complete biosynthetic pathways or even complete genomes will be an important enabling technology for synthetic biology. Here we describe a rapid, iterative method based on site-specific lambda Red recombination for assembling fosmid clones into larger episomal constructs. Fosmid clones are convenient for this purpose because the fosmid library construction procedure is routine and, at approximately 40 kbp, fosmid clones are large enough to contain entire bacterial operons and gene clusters but are still small enough to be easily sequenced and manipulated. A 40 kbp DNA segment is within the scope of what can be synthesized in vitro and, therefore, there is a convenient transition from in vitro construction of fosmid sized DNA fragments to in vivo assembly of these precursors into substantially larger DNA molecules.

We are exploring approaches for building microbial chromosomes (Holt et al. 2007). Recently, it has been shown that the entire 1.1 Mbp chromosome from *Mycoplasma mycoides* can be purified by PFGE and transferred into a recipient cell of the closely related species, *Mycoplasma capricolum* (Lartigue et al. 2007). The transfer event was facilitated by polyethylene glycol, a reagent typically used to promote mammalian cell fusion in vitro. *Mycoplasmas* are amenable to this approach because they lack a cell wall. Selection for markers on the *M. mycoides* chromosome allowed isolation of cells identical to this species and harboring only this genome. This is an important advance that shows that it is possible to exchange *Mycoplasma* chromosomes. However, approaches for actually building microbial genome require further development and, since there are no current in vitro synthesis methods to construct a DNAs that are 100s of kbp or larger, the most practical approach to building a genome, be it natural, modified, or fully synthetic, appears to be stepwise assembly of moderately large segments in a host cell. As such, there must be co-residence of the donor and host genomes in a single cell for some period of time as the

construction procedure moves to completion, and a means to recognize and mitigate incompatibilities between donor and host gene products. Regarding the methodology for actually constructing a microbial genome in a host cell, various approaches can be envisioned, and some progress has already been made. Itaya et al. (2005) used an iterative process of homologous recombination to integrate most (3.5 Mbp) of the genome of the photosynthetic bacterium *Synechocystis* directly into the genome of a *Bacillus subtilis* host. Their process, termed “inchworm elongation”, involves insertion of a target sequence into the *Bacillus* genome followed by delivery of a *Synechocystis* genome segment tens of kilobases in length that recombines at that site. This is done iteratively, in an “inchworm” fashion in order to establish longer donor segments within the host chromosome and shows that large segments of a bacterial genome can be assembled in a host cell. Our approach to fosmid clone assembly that is described in the present study requires relatively few steps, and has the advantage that the DNA molecule is constructed as an episome, which can be isolated from the host for further analysis, manipulation or transplantation. Mechanistically, our approach is similar to that described by Kotzamanis and Huxley (2004). This group demonstrated the fusion of a pair of overlapping human BAC clones by first subcloning inserts into new vectors that carried different antibiotic resistance genes, then joining them by lambda RED recombination with antibiotic switching. Our study extends this work by establishing a fosmid vector system for constructing recombination-ready libraries, and demonstrating serial rather than just pairwise recombination. Further, we have successfully mitigated the issue of executing recombination in the presence of a highly similar host genome. Further, we anticipate that hierarchical assembly involving iterative steps of pairing fosmids, then joining these pairs, and so on, has the potential to dramatically accelerate the assembly of very large DNAs.

A difficulty that can be anticipated in assembling large episomal elements in closely related host cell is that some number of genes encoded by a given segment of DNA will be transcribed and translated. This ectopic expression may have a deleterious effect on the host through mechanisms that may include, for example, direct toxicity of gene products, altered gene dosage, or sequestration of rare codons. As episomal elements are assembled, there is increasing chance of reconstituting a set of genes that are

individually tolerable, but toxic in combination. The study of the combinatorial behavior of gene products is an important area of synthetic biology, yet remains underexplored. The method of iterative clone recombination presented here will facilitate these studies, and provides a useful and broadly applicable approach to building large DNAs.

**Acknowledgements** R.A.H. is a Michael Smith Foundation for Health Research Scholar. We thank Dr. Rosie Redfield (University of British Columbia) for *H. influenzae* Rd KW20, and Dr. Court (NCI-Frederick) for *E. coli* EL350 cells. Funding for this work was provided by Genome British Columbia.

## References

- Carlson R (2003) The pace and proliferation of biological technologies. *Biosecr Bioterror* 1:203–214
- Cello J, Paul AV, Wimmer E (2002) Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of a natural template. *Science* 297:1016–1018
- Holt RA, Warren R, Flibotte S, Missirlis PI, Smailus DE (2007) Rebuilding microbial genomes. *Bioessays* 29:580–590
- Huang X, Wang J, Aluru S, Yang S, Hillier L (2003) PCAP: a whole-genome assembly program. *Genome Res* 13:2164–2170
- Itaya M, Tsuge K, Koizumi M, Fujita K (2005) Combining two genomes in one cell: stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc Natl Acad Sci USA* 102:15971–15976
- Kotzamanis G, Huxley C (2004) Recombining overlapping BACs into a single larger BAC. *BMC Biotechnol* 4:1
- Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC (2007) Genome transplantation in bacteria: changing one species to another. *Science* 317:632–638
- Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, Court DL, Jenkins NA, Copeland NG (2001) A highly efficient *Escherichia coli*-based chromosome engineering. *Genomics* 1:56–65
- Smailus DE, Marziali A, Dextras P, Marra MA, Holt RA (2005) Simple, robust methods for high-throughput nanoliter-scale DNA sequencing. *Genome Res* 15:1447–1450
- Smith HO, Hutchison CA III, Pfannkoch C, Venter JC (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc Natl Acad Sci USA* 100:15440–15445
- Tian J, Gong H, Sheng N, Zhou X, Gulari E, Gao X, Church G (2004) Accurate multiplex gene synthesis from programmable DNA chips. *Nature* 432:1050–1054
- Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci USA* 97:5978–5983