

Microbial Genome Reconstruction

by Iterative Clone Recombination

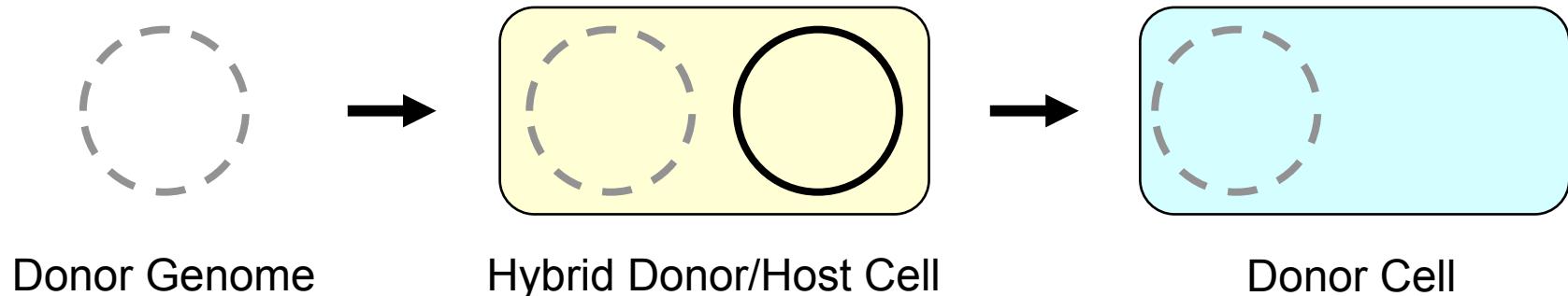
Synthetic Biology 3.0
Zurich, Switzerland
René L. Warren - June 25th 2007



Genome Reconstruction

Using a top-down approach

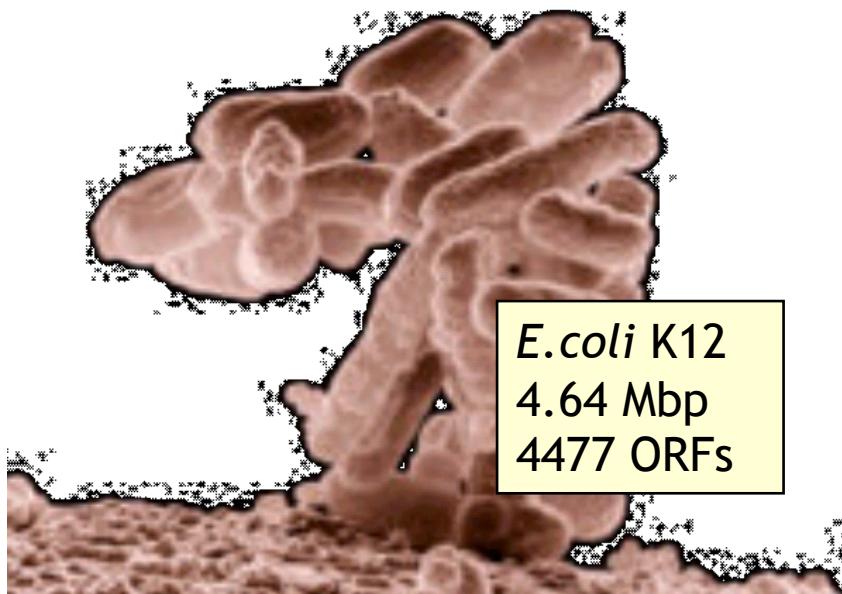
- Akin to reverse engineering
- Uses known blueprints & tools to rebuild a genome
- Uses a host to activate ‘naked DNA’ & obtain the organism it encodes



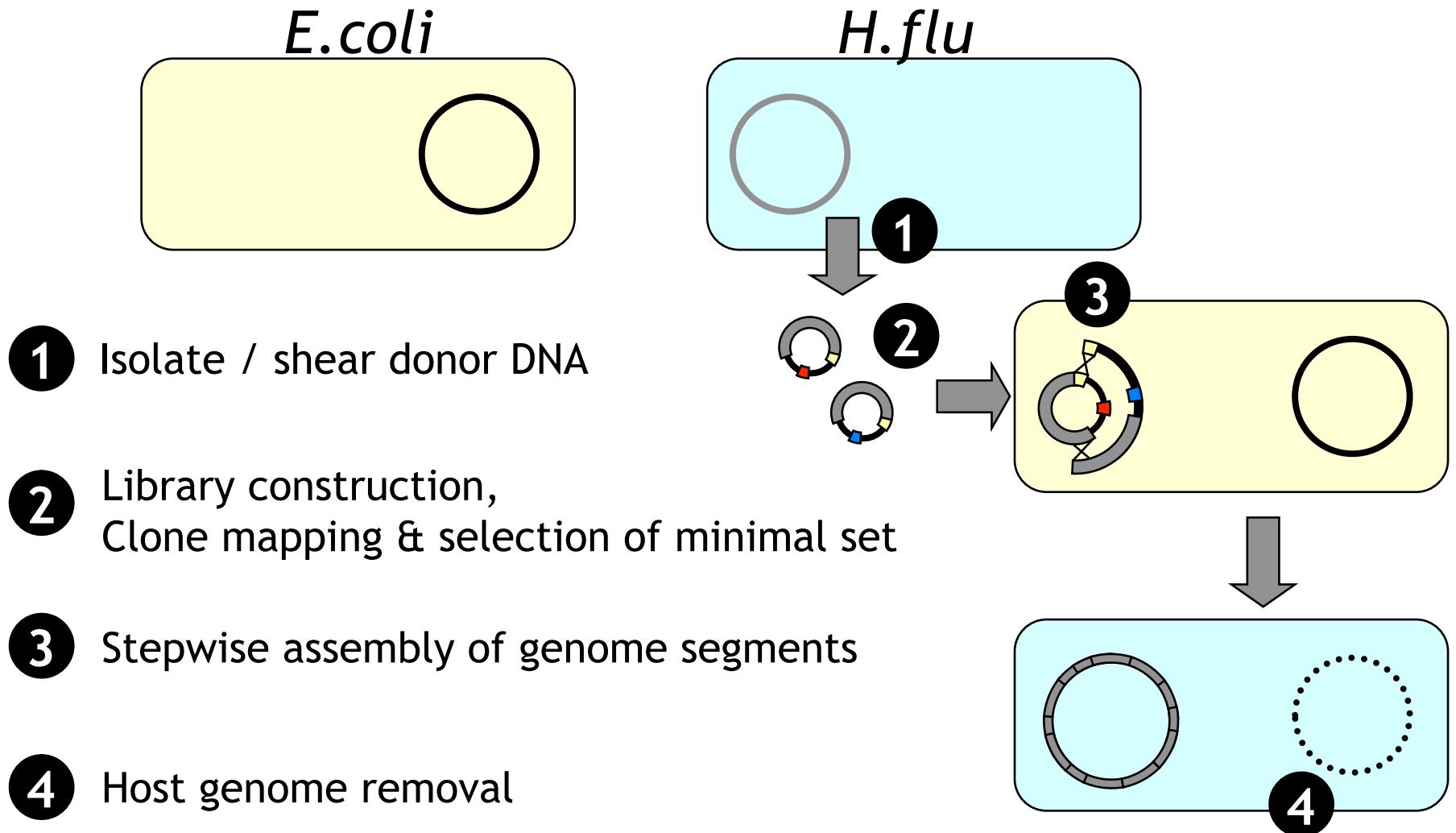
Model

Using *E.coli* to rebuild *H.influenzae* (*H.flu*)

- Both gram-negative commensal gamma proteobacteria
Favors compatibility of components
- Well-characterized, non-pathogenic free-living lab strains
Genome sequences are known
- *H.flu* genome is relatively small
Minimize number of manipulations

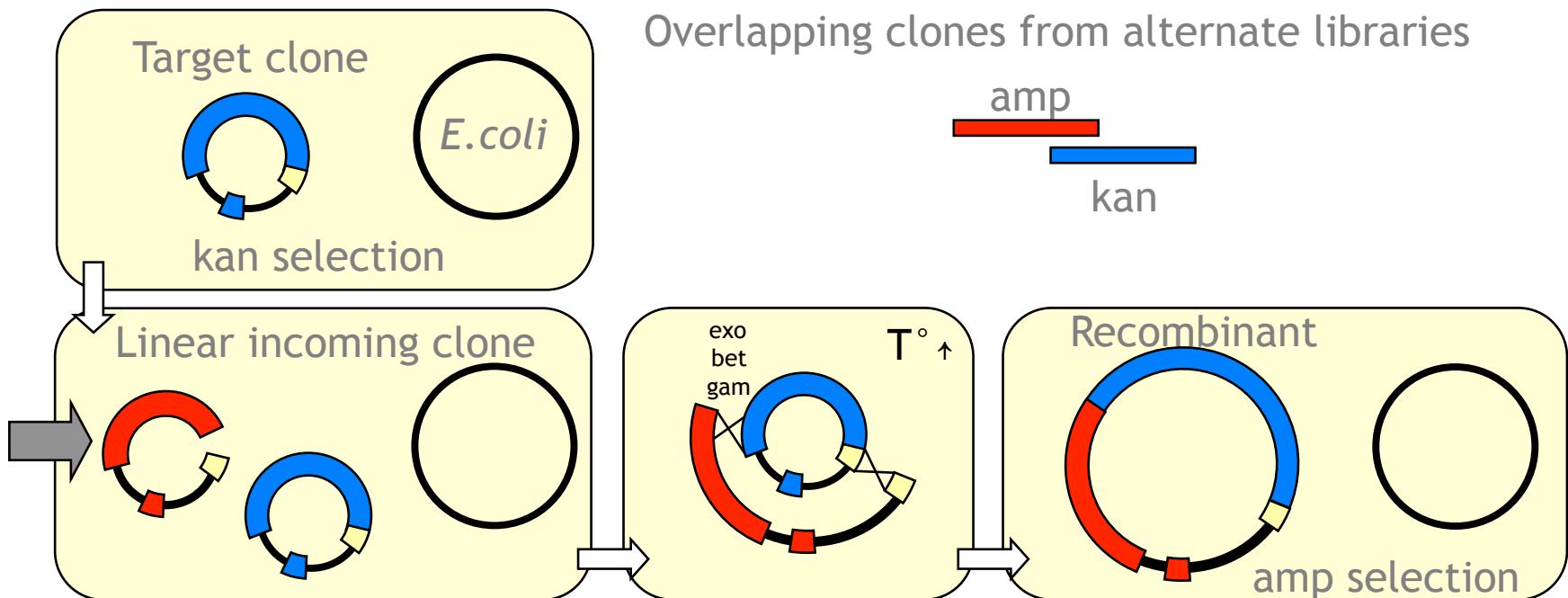


Top-down in a Nutshell



Recombination Strategy

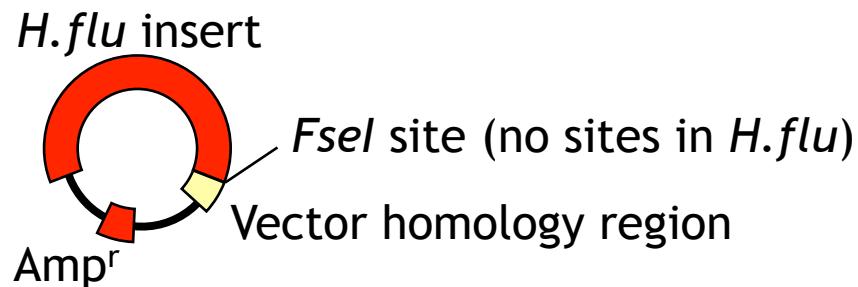
- Lambda Red recombination (Yu et al. PNAS 2000)
Host carries segment of the phage λ genome: *exo*, *bet* & *gam* genes
- Genes mediate *in vivo* recombination between ends of a linear incoming DNA fragment with homologous sequences in a target.
- Induced by temperature.
- Overlap can be short (~50 bp)
- Linear incoming clone can't propagate in the host, allowing for selection of recombinants
- Recombination is seamless



Genomic Libraries

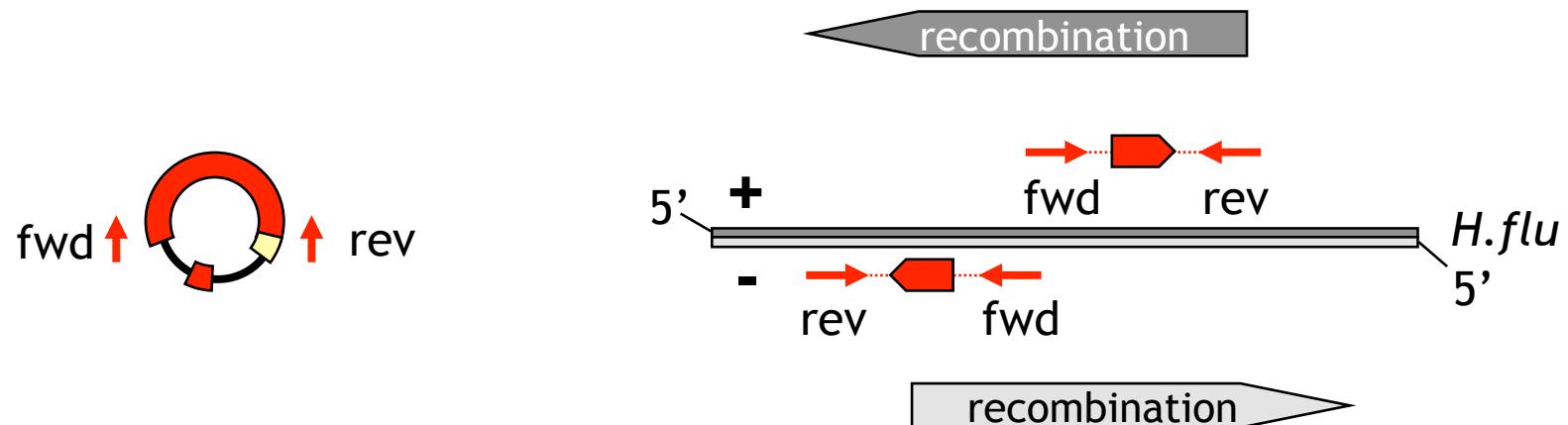
- Two fosmid libraries constructed
 - Kanamycin (Kan^r) / Ampicillin (Amp^r) resistance for selecting recombining clones
 - ~5000 fosmid clones end-sequenced, 2500 from each library
- Cloning vectors engineered for lambda Red recombination
 - Fosmids chosen as vectors
 - Easier to manipulate than BACs
 - Finer resolution (smaller inserts) supports skipping over genomic regions if needed

Unique restriction site to linearize incoming clone prior to recombination



Clone Mapping

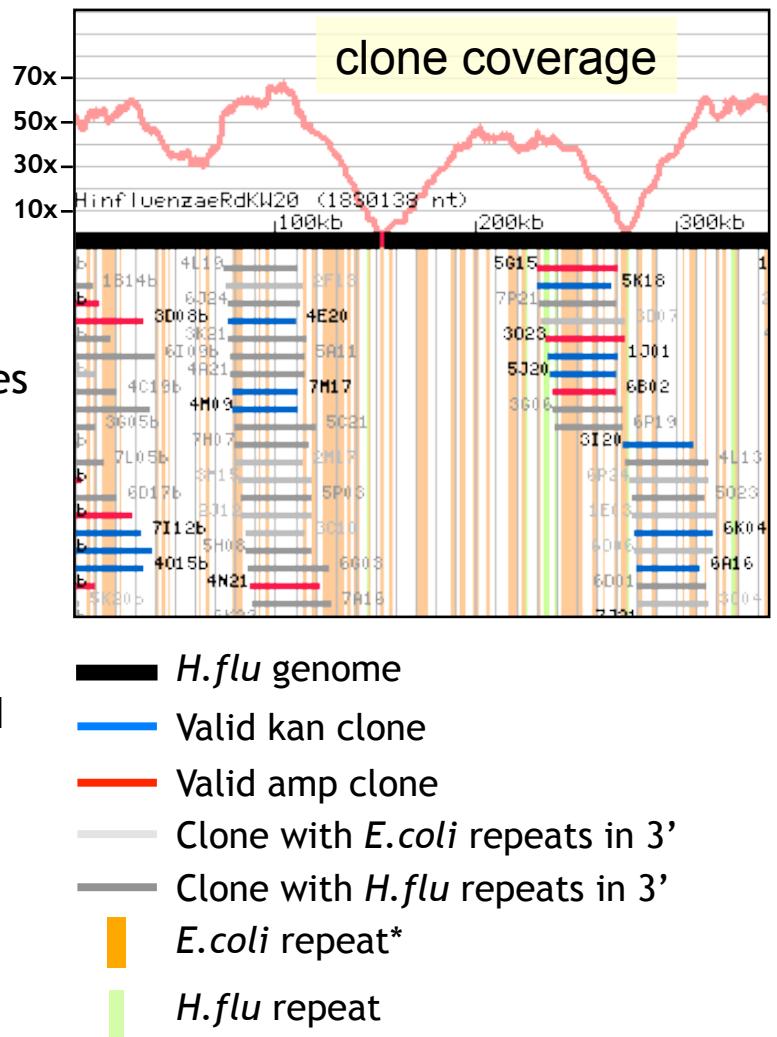
- Fosmid clones prep'd and sequenced on ABI3730XL
- Low quality/vector bases trimmed
- Paired-end sequences mapped onto *H.flu* genome
 - Pairing logic preserved
 - Pairs align in opposite direction, facing inwards
 - Clone size insert size respected (~40kb ± 2 std.dev.)
 - Selection is directional



Clone Selection

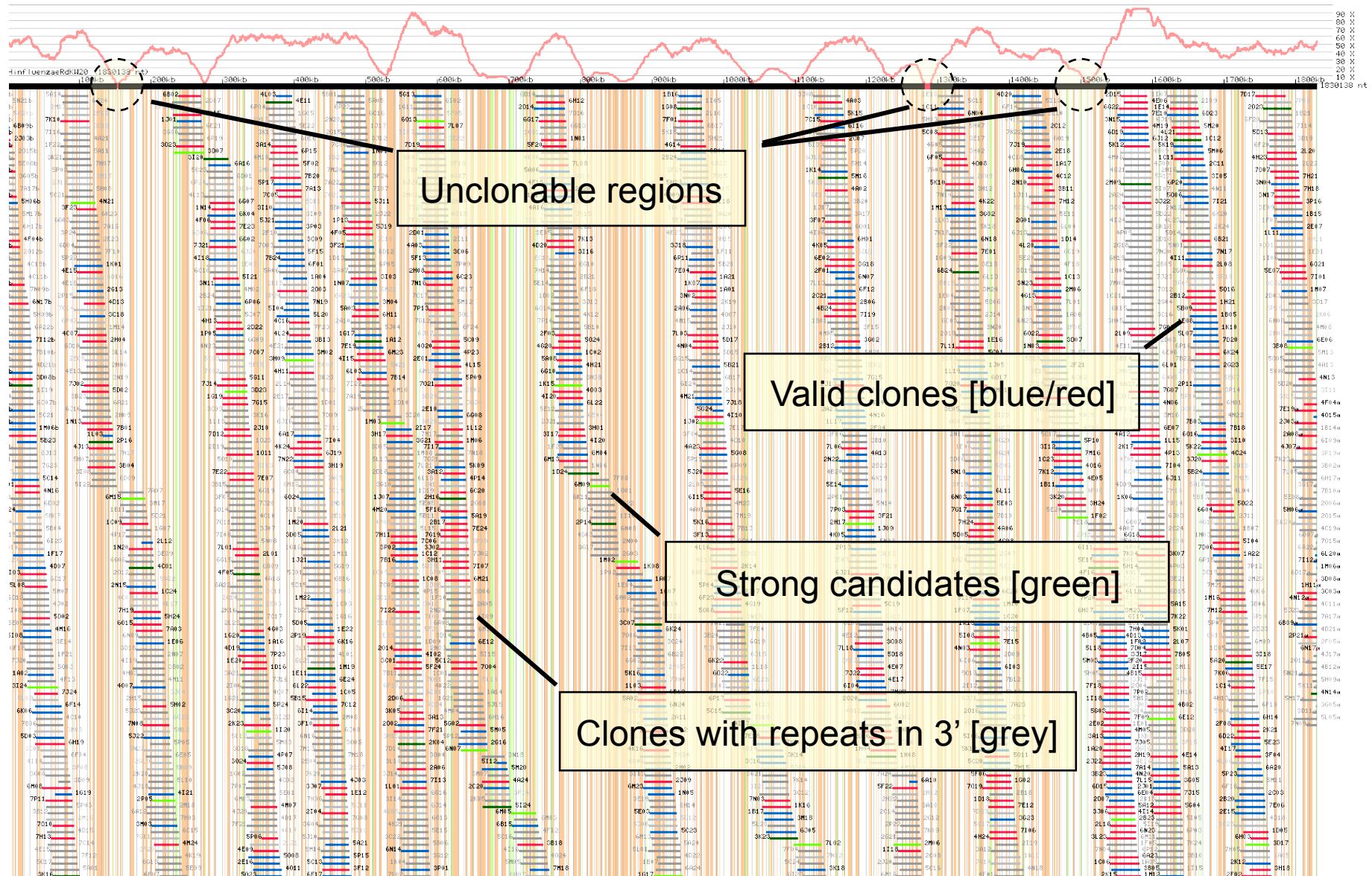
Favor largest clones:

- From strand with highest coverage
40x average clone coverage on plus strand
~2000 logical clones covering 99.21% of genome
- From alternate libraries (kan-amp)
Key to recombination strategy - allow screening of recombinants
- 3'-end (~50nt) of the 5'-most recombining clone does not align to *E.coli* nor with *H.flu* outside intended region
No recombining end
- Overlap minimally with neighbors
Favor those > 0.5kb and < 10kb
- Software developed to map and visualize clones and repeats
- Selection performed in a semi-automated fashion
Suitable clones are flagged by the software
- Final selection checked manually
For optimum overlap and confirm absence of repeats



* Showing repeats $\geq 20\text{nt}$ & $\geq 70\%$ seq. id.

Clone Selection

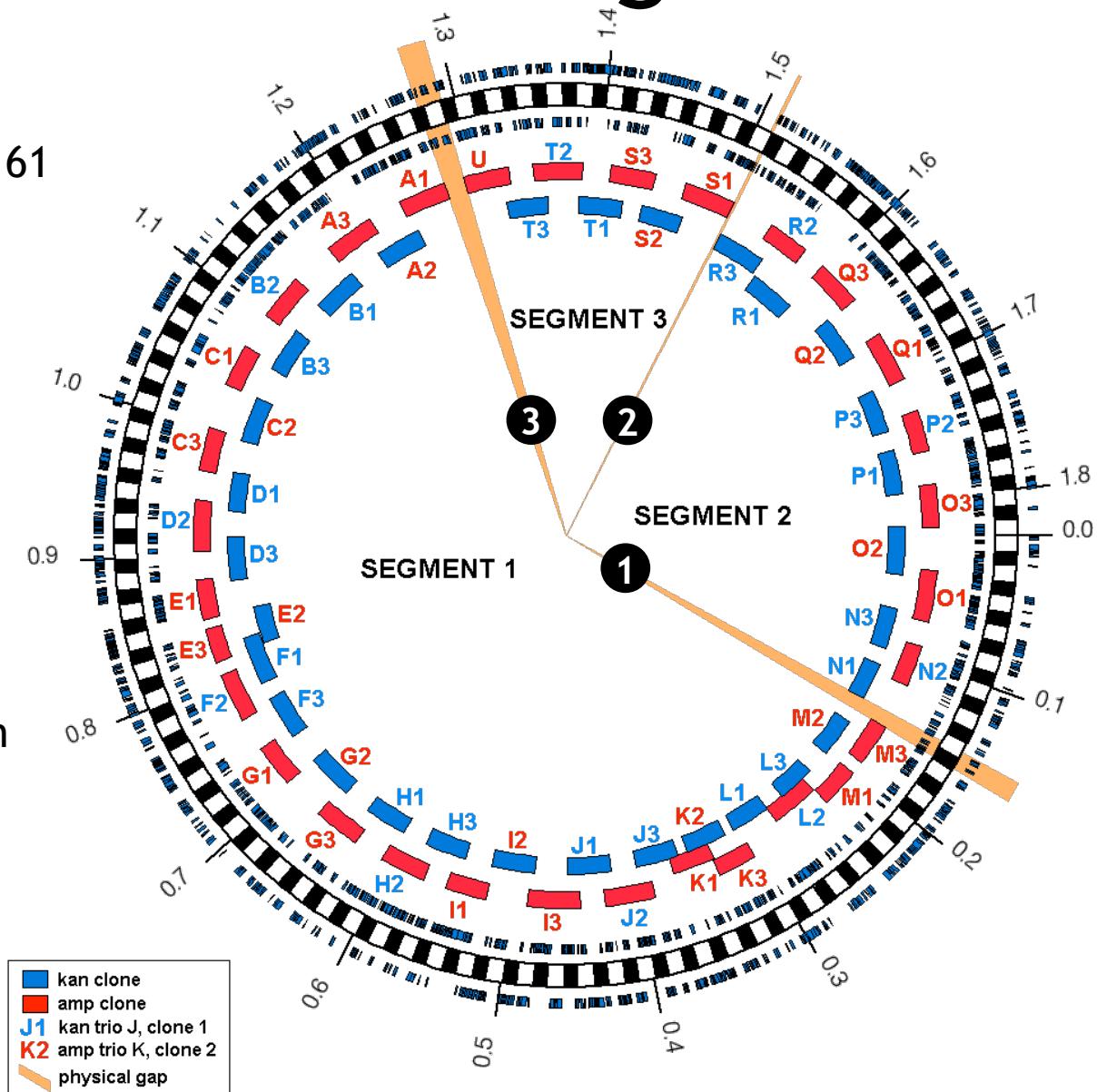


H. flu Minimum Tiling Path

- Tiling path consist of 61 alternating clones
30 kan
31 amp

- Three gaps in path
 - 1 11.4 kb
 - 2 1.7 kb
 - 3 15.3 kb

These gaps contain
29 fully contained
or straddled ORFs



What's Missing ?

Gap1 - 11.4kb

<i>afuA</i>	ferric ABC transporter protein
<i>udk</i>	uridine kinase
<i>dcd</i>	dCTP deaminase
<i>HI0134</i>	hypothetical protein
<i>HI0135</i>	conserved hypothetical protein
<i>engA</i>	GTP-binding protein
<i>dnaQ</i>	DNA polymerase III, epsilon subunit
<i>rnhA</i>	ribonuclease H
<i>ompP2</i>	outer membrane protein P2
<i>nagA</i>	N-acetylglucosamine-6-phosphate deacetylase
<i>nagB</i>	glucosamine-6-phosphate isomerase
<i>aspS</i>	aspartyl tRNA synthase

Gap2 - 1.7kb

HI1418	phage antirepressor
--------	---------------------

Gap3 - 15.3kb

<i>argR</i>	arginine repressor
<i>mdh</i>	malate dehydrogenase
<i>lysS</i>	lysyl-tRNA synthase
<i>prfB</i>	peptide chain release factor 2
<i>dsbC</i>	disulfide interchange protein
<i>recJ</i>	ss exonuclease
<i>dsbA</i>	disulfide interchange protein
<i>lctP</i>	L-lactate permease
<i>cmkA</i>	cytidylate kinase
<i>rpsA</i>	ribosomal protein S1

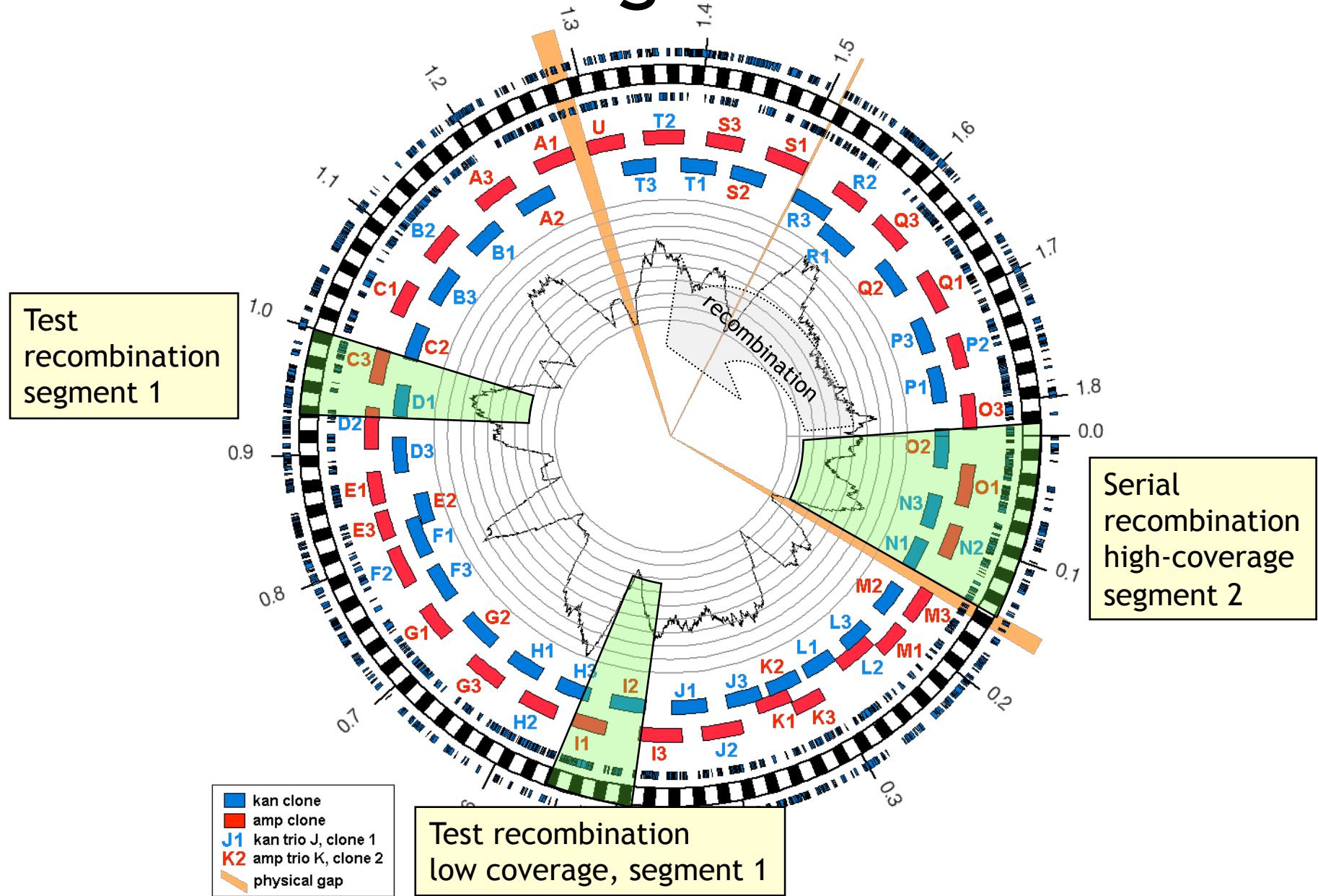
H. flu genes that could not sustain transposon insertion (Akerley et al. 2002)

Part of the essential core of a minimal gene set required for cell viability (Gil et al. 2004)

- Meta-analysis of 8 bacterial WGA identifies recurring genes in clone gaps (Holt et al. Bioessay 2007)
-e.g. ribosomal proteins, disulfide isomerase, tRNA synthase-coding genes

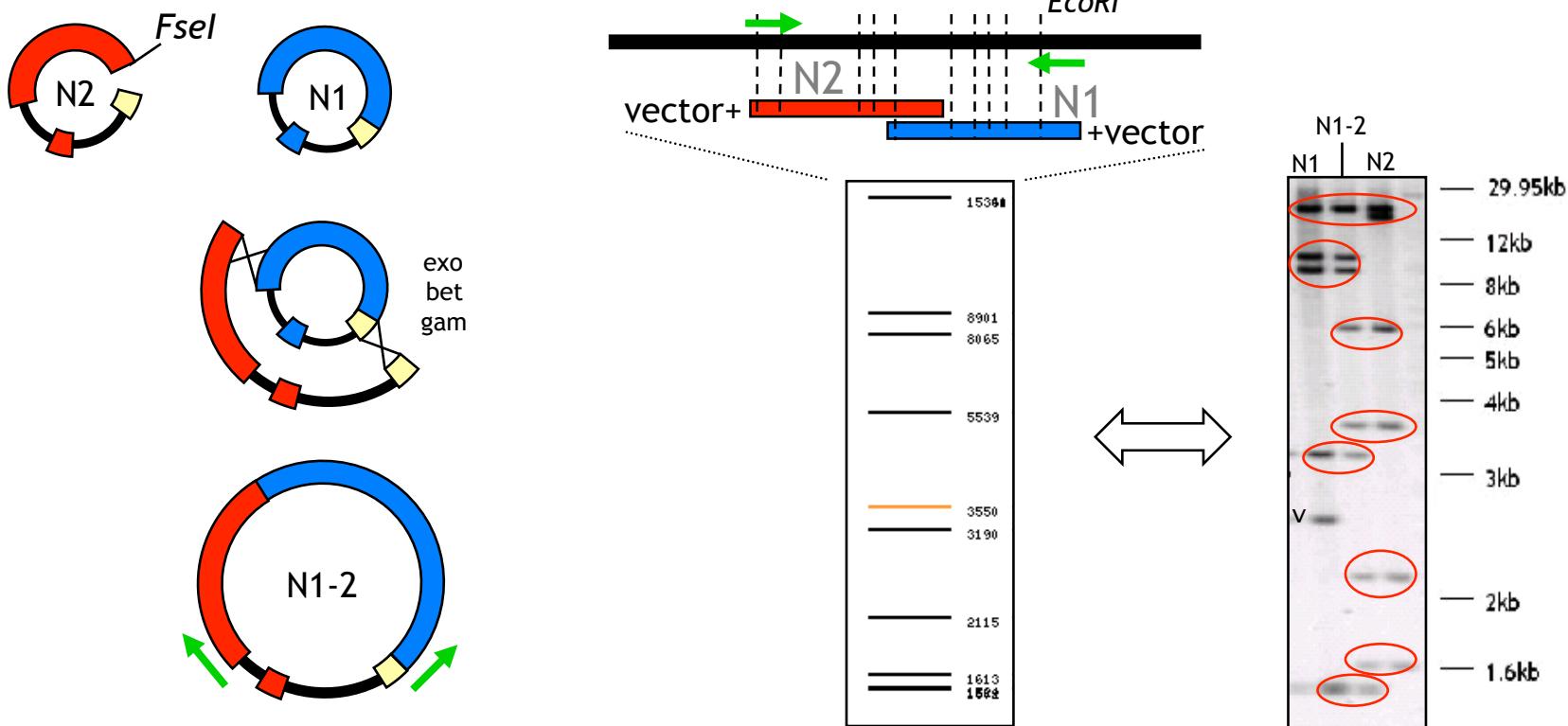
See poster: Holt RA et al “A Model System for Rebuilding Microbial Genome”
presented tonight

Progress

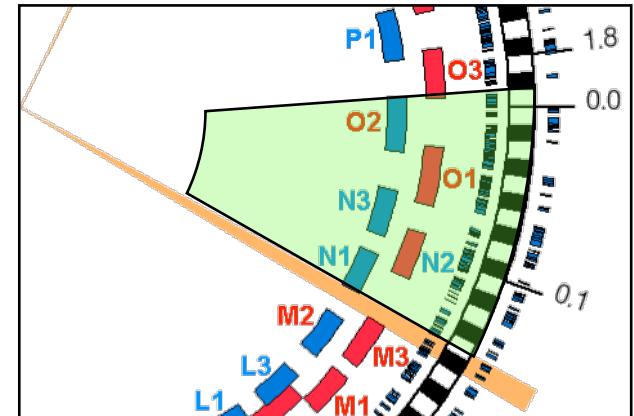


Monitoring Recombinants

- End-sequencing and mapping sequence reads onto the *H. flu* genome
 - gives estimates on recombinant size
- Fingerprinting and comparison to *in-silico* digests
- Pulse-field gel electrophoresis (recombinant size)
- PCR-amplification of overlap junction (if overlap permits - not shown)



Results



PFGE IMAGE HERE

High-Throughput Strategies

- High-throughput strategies designed to
 - Reduce manipulations
 - Increase speed

1. The “Trio” strategy allow parallel recombinations

2. The “Shotgun” strategy relies on cell pooling to reduce number of manipulations

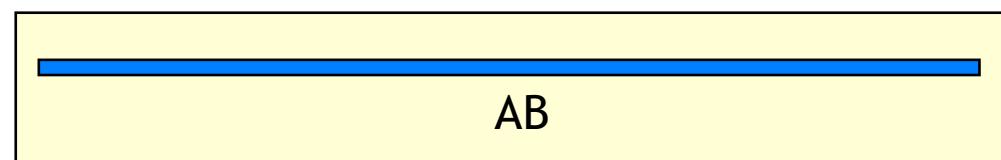
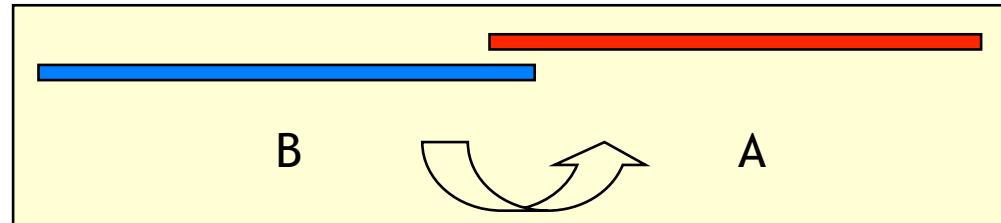
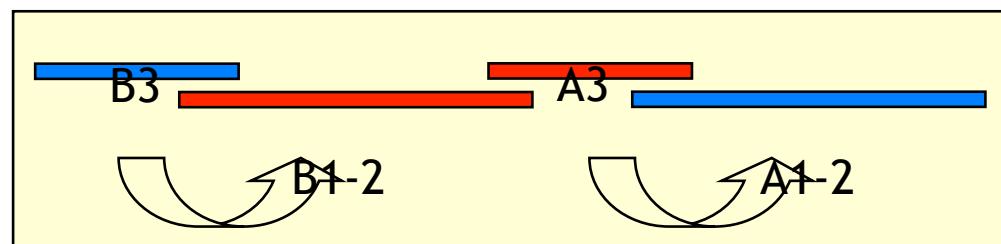
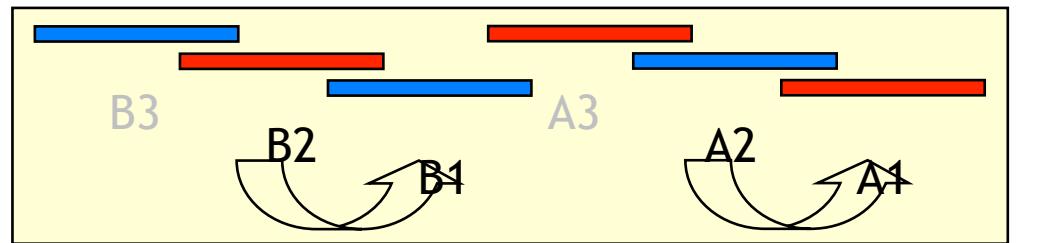
e.g.

all n_1 amp targets are pooled
all n_2 kan incomings are pooled
clones are recombined at once

(not shown)

H. flu
genome

recombination



The “trio” strategy

Conclusion

- Developed a top-down approach for rebuilding a bacterial genome in a host
 - Gap in biological knowledge makes reverse-engineering attractive
 - Uses host to rebuild donor genome
 - Uses host components to activate donor DNA
 - Host chosen for compatibility with donor
- 8.4% of *H.flu* genome rebuilt in *E.coli*
 - Segment rebuilt using fosmids -- more granular than BACs
 - Dual fosmid library system designed for selection of recombinants
 - Lambda Red recombination is efficient and seamless
 - Segment rebuilt so far propagates well in the host
- Devised high-throughput recombination strategies
 - Speed up the recombination process
 - Reduce manipulations

Outstanding Issues

- Genome Activation
 - Currently monitoring *H.flu* gene expression and host response
 - Strain compatibility crucial in activating donor genome
- Delivering essential, yet toxic genes & reconstituting metabolic pathways toxic to *E.coli*
 - Envision clone “by-pass” - allowed due to resolution of fosmids
 - Skipped clones/segments would be added at the end as a simple construct to facilitate genome resolution
- Genome resolution
 - Disrupt host partitioning system
 - Host genome ablation using restriction enzymes
 - Curing by antibiotic selection or inducible “suicide” gene
 - Other approaches can be envisioned, but not tested until assembly of donor genome is complete

Acknowledgments

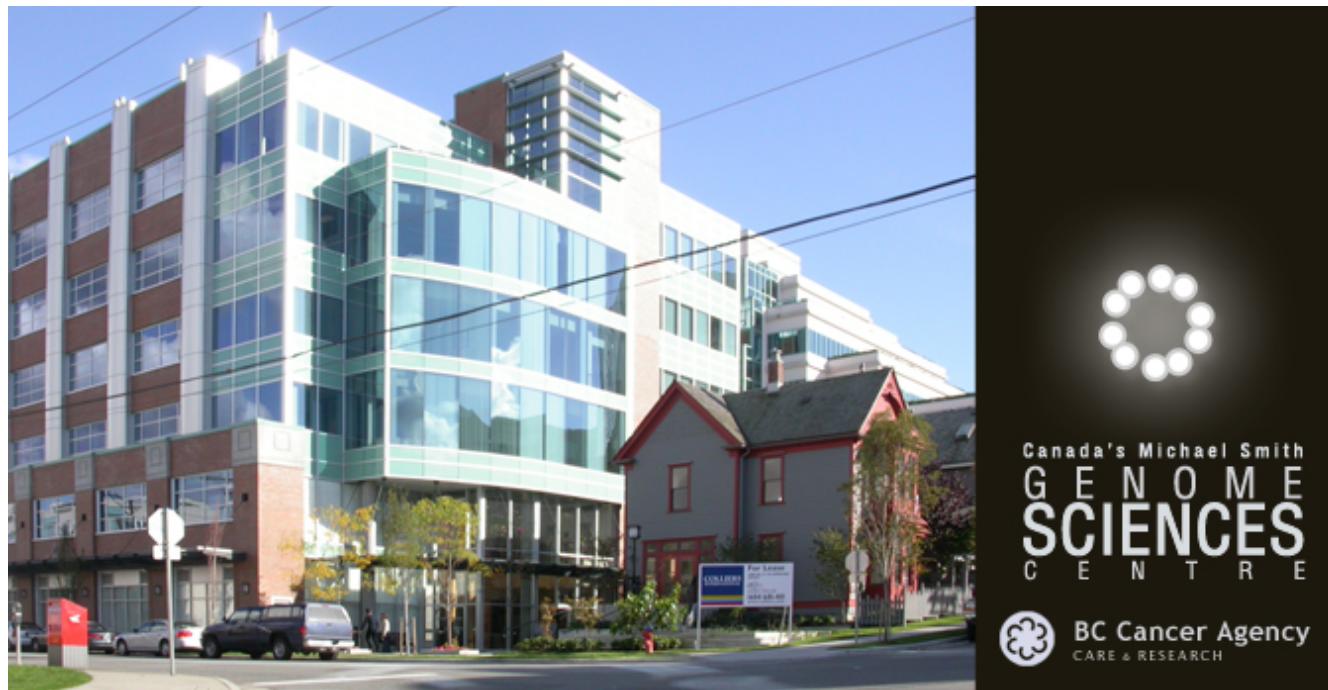
Funds provided by:

Microsoft Research

Genome British Columbia

BC Cancer Agency

- Rob Holt*, PI
- Duane Smailus, laboratory
- Doug Freeman, laboratory
- Perseus Missirlis, informatics



*R.Holt is a Michael Smith Foundation for Medical Research Scholar

