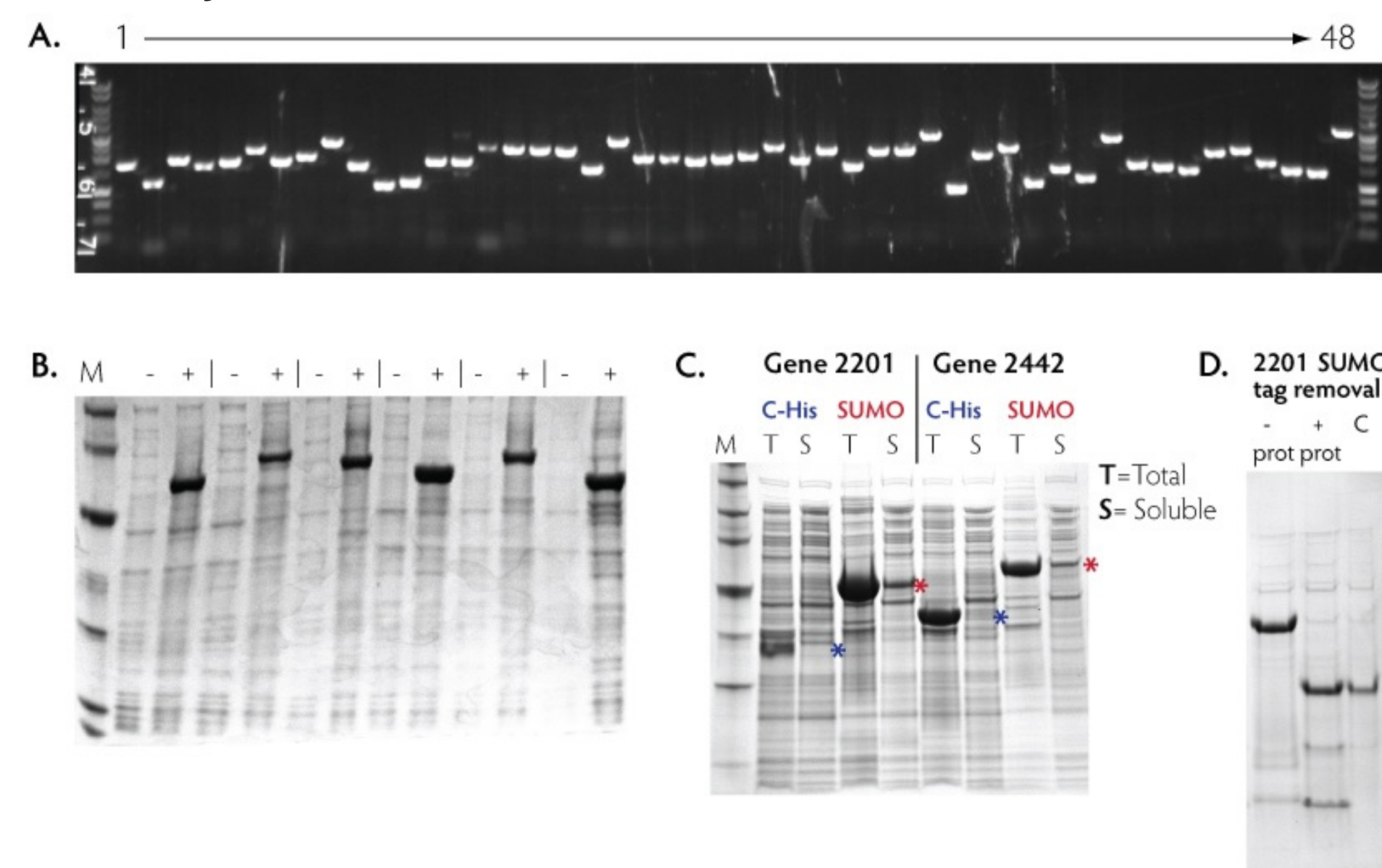


ABSTRACT

The efficient capture and functional analysis of genes and metabolic pathways is constrained by the choice of available tools. A directed cloning system has been developed that uses *in vivo* homologous recombination to seamlessly join PCR amplified genes with pre-processed plasmids, eliminating numerous time consuming and expensive reagents and steps. Unlike other cloning systems, no vector preparation, restriction or modifying enzymes, or purification steps are required. The desired insert is simply amplified with primers that include 18 bases of overlap with the ends of a small expression vector, and mixed directly with the plasmid preparation and competent cells. >90% of clones have the target gene inserted in the correct orientation. Using a multiplex assay for *endo* and *exo*-cellulases and hemicellulases in a microplate format, we were able to efficiently screen thousands of CAZymes/week without robotics. We have demonstrated the utility of the system by cloning, expression, purification and characterization of over a hundred different CAZymes from thermophilic, mesophilic and alkaliphilic microbes including *Dictyoglomus turgidum*, *Acidothermus cellulolyticus*, *Fibrobacter succinogenes*, and *Bacillus cellulolyticus*.

A second tool has been developed that can clone and express ~ 30 kb segments of random DNA or large metabolic pathways. Cloning and expression of large operons with conventional vectors presents numerous challenges due to: 1) Secondary structure associated with repetitive sequences or high AT-content, which can lead to DNA deletions; 2) Unregulated transcription that can select against toxic, but potentially valuable genes; 3) High copy number vectors and gene dosage that creates a large metabolic burden on the cell; 4) Strong promoters that can cause overexpression of deleterious proteins; 5) Gene instability due to superhelical stress seen in large insert circular plasmids; and 6) General decrease in cloning efficiency with larger fragments. We have developed a new pathway capture and engineering tool to clone and express multiple genes from large pathways. A linear "pJAZZ" vector for *E. coli* expression readily holds large inserts (up to 30 kb) of DNA. Its low copy number (2-4/cell) is ideal for high stability protein production. Arabinose-inducible amplification of copy number (5-20 fold) further improves cloning of difficult sequences. The system has been validated for expression of the nine gene MEP isoprene pathway. We have also found functionally active thermostable DNA polymerases and their flanking accessory proteins from hot spring viral metagenomic samples with this new tool.

Figure 3. Expressioneering large-scale cloning and expression case study.



(A) PCR products from 48 putative *Fibrobacter succinogenes* hydrolase genes ranging from ~1 to > 3 kb. These PCR products were cloned into the pETite C-His vector. (B) Uninduced (-) and IPTG-induced (+) samples of HI-Control BL21(DE3) Cells with 6 different genes cloned into the pETite C-HIS Vector. (C) Enhanced solubility of SUMO-tagged 2201 and 2442 gene products. Total cell extract and soluble fractions are shown. (D) Removal of 6xHis-SUMO tag from purified SUMO-2201 fusion protein by SUMO protease. -prot: uncleaved SUMO-2201 fusion protein after IMAC purification; +prot: SUMO protease-treated fusion protein; C: isolated 2201 protein after removal of 6xHis-SUMO fragment and SUMO protease by subtractive IMAC.

Figure 1. Expressioneering vectors pETite & pRham designed for instant enzyme-free cloning of PCR products.

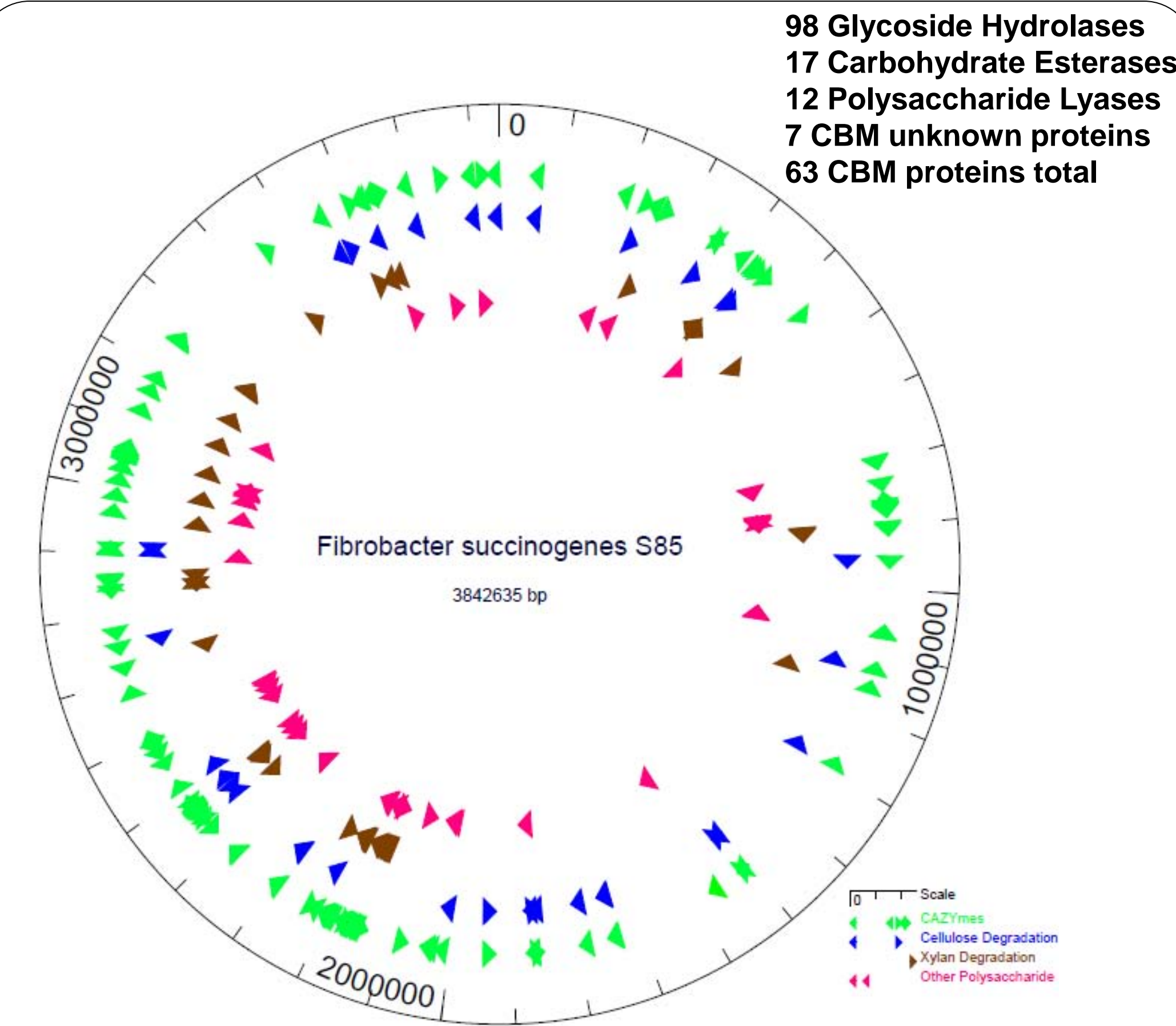
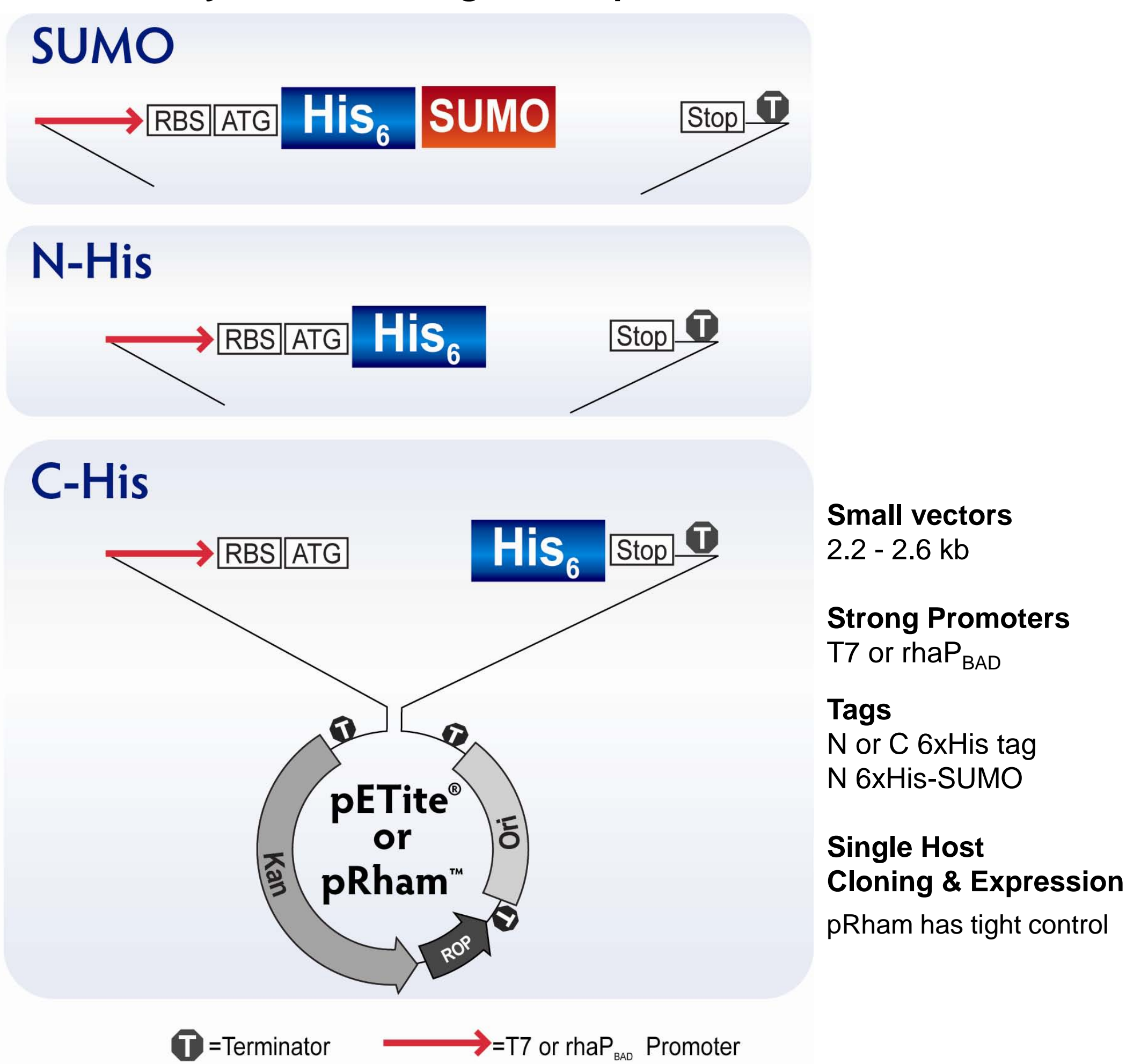
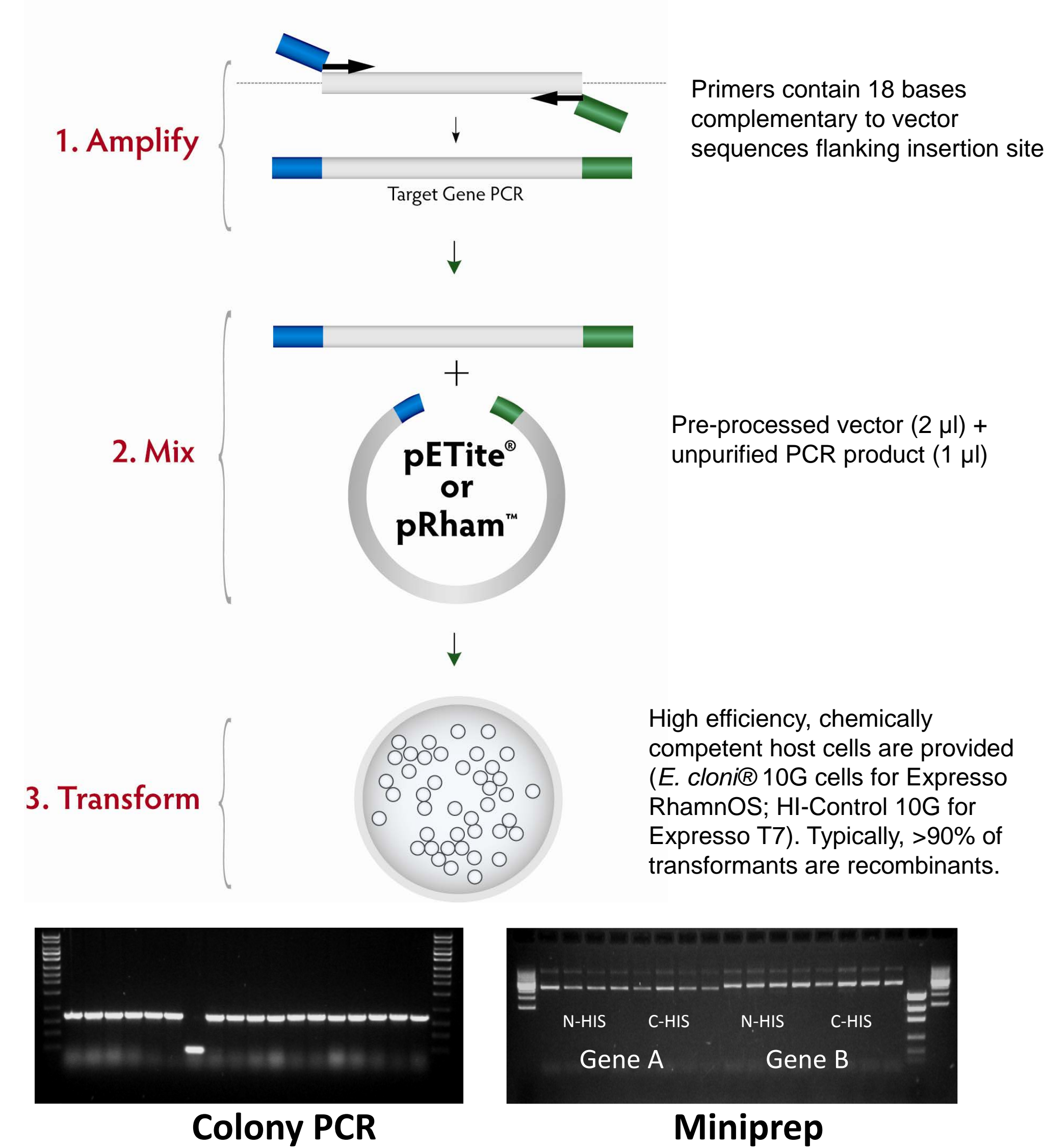


Figure 4. Exploring the *Fibrobacter succinogenes* GH genome. Functional Analysis of 100 CAZY Genes

Suen et al. 2011. The Complete Genome Sequence of *Fibrobacter succinogenes* S85 Reveals a Cellulolytic and Metabolic Specialist. PLoS ONE 6(4):e18814.

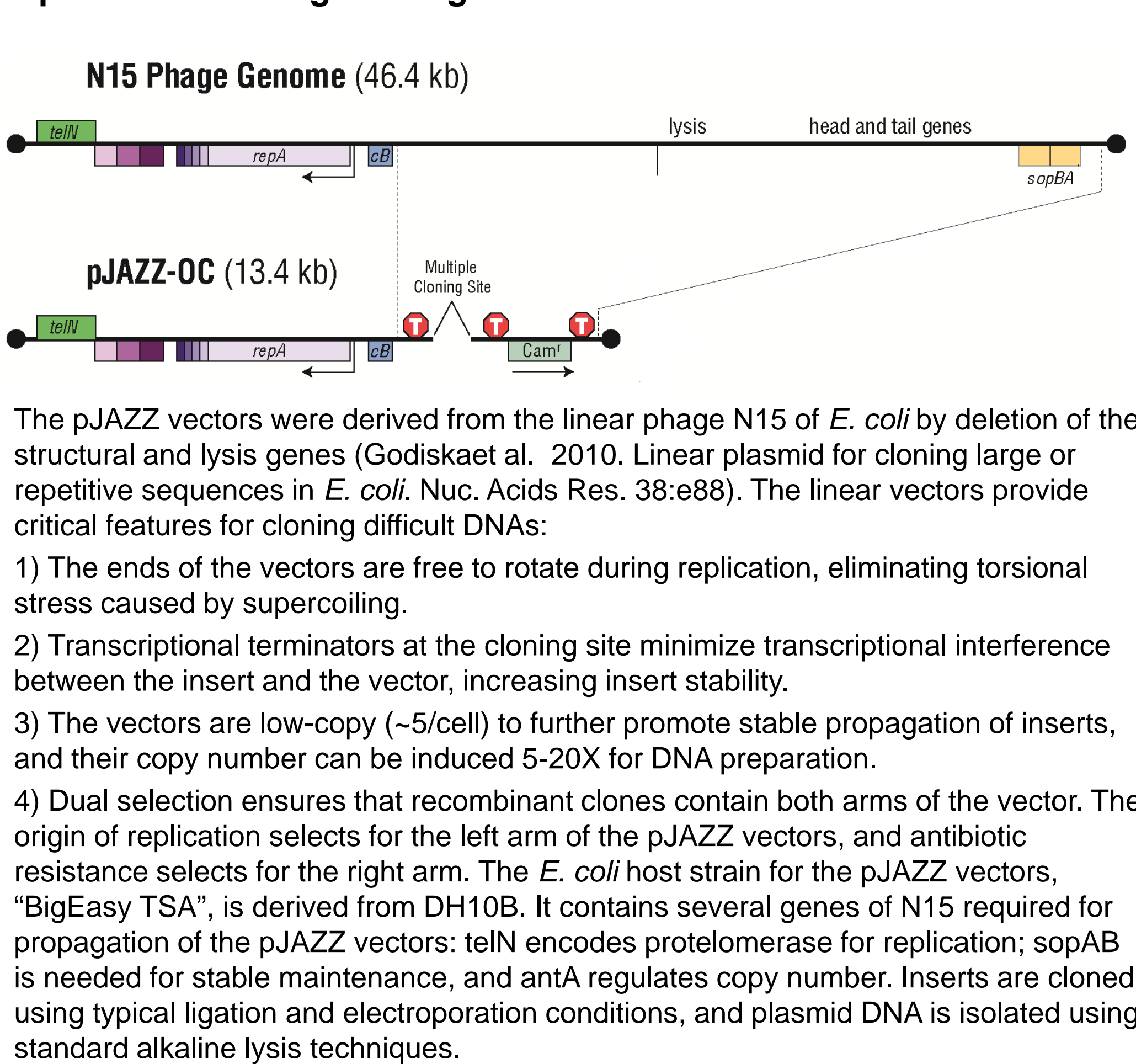
Figure 2. Five second enzyme-free expression cloning.

Steinmetz (2011) Expresso® Cloning and Expression Systems: Expressioneering™ Technology streamlines recombinant protein expression. Nature Methods 8(6) p. iii-iv



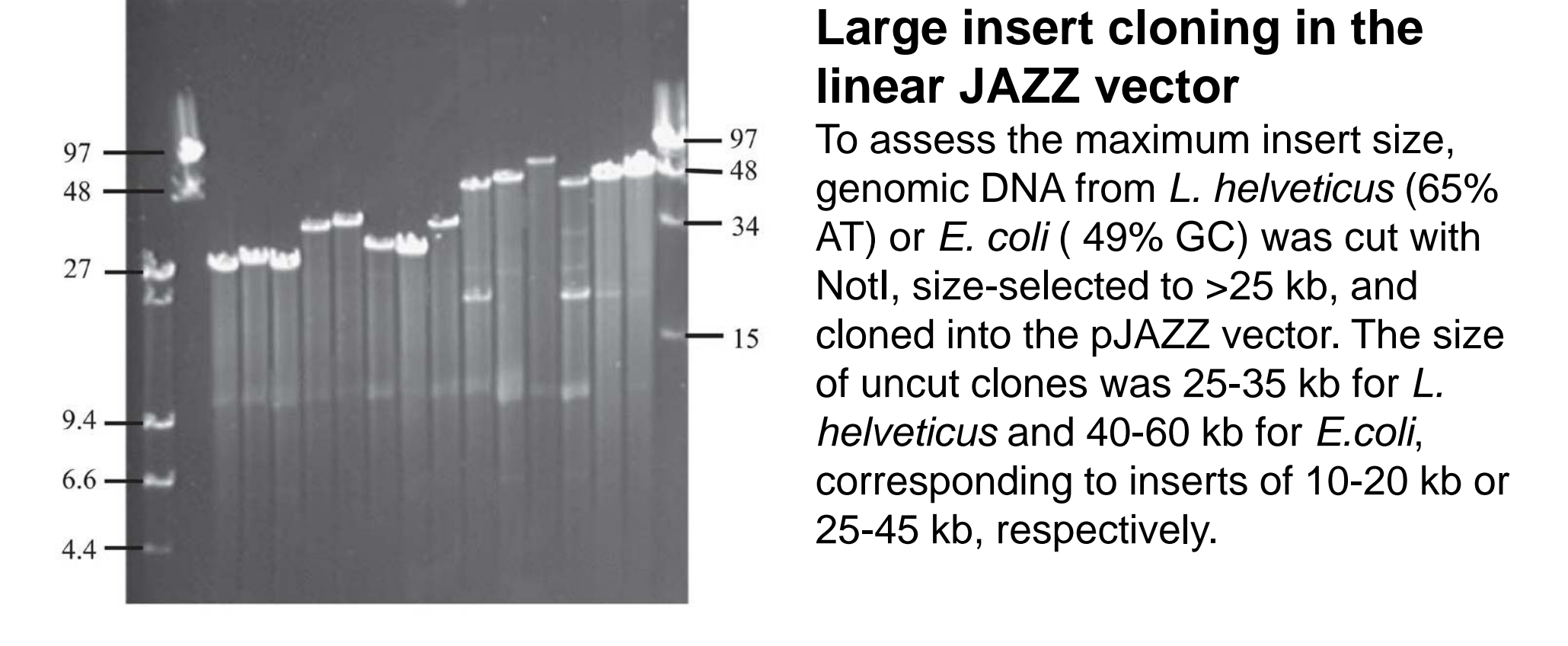
The high efficiency of the recombination-based cloning strategy means minimal screening is necessary to identify positive clones. Primers for clone validation by PCR and sequencing are provided in the Expresso kits. In the left panel above, colony PCR was performed on 18 candidate clones of a potentially toxic 0.55 kb gene in the pETite C-His Vector; all but one contained insert of the correct size. Sequencing verified that the remaining 17 clones were correct. In the right panel, 2 different genes were cloned into both the pETite N-His and pETite C-His vectors. Plasmid DNA was isolated by miniprep from four randomly-selected colonies of each transformation plate; all contained inserts of correct size.

Figure 5. Novel linear phage derivative for large insert pathway expression and engineering in *E. coli*.



The pJAZZ vectors were derived from the linear phage N15 of *E. coli* by deletion of the structural and lysis genes (Godiska et al. 2010. Linear plasmid for cloning large or repetitive sequences in *E. coli*. Nuc. Acids Res. 38:e88). The linear vectors provide critical features for cloning difficult DNAs:

- 1) The ends of the vectors are free to rotate during replication, eliminating torsional stress caused by supercoiling.
- 2) Transcriptional terminators at the cloning site minimize transcriptional interference between the insert and the vector, increasing insert stability.
- 3) The vectors are low-copy (~5/cell) to further promote stable propagation of inserts, and their copy number can be induced 5-20X for DNA preparation.
- 4) Dual selection ensures that recombinant clones contain both arms of the vector. The origin of replication selects for the left arm of the pJAZZ vectors, and antibiotic resistance selects for the right arm. The *E. coli* host strain for the pJAZZ vectors, "BigEasy TSA", is derived from DH10B. It contains several genes of N15 required for propagation of the pJAZZ vectors: telN encodes protelomerase for replication; sopAB is needed for stable maintenance, and antA regulates copy number. Inserts are cloned using typical ligation and electroporation conditions, and plasmid DNA is isolated using standard alkaline lysis techniques.



To assess the maximum insert size, genomic DNA from *L. helveticus* (65% AT) or *E. coli* (49% GC) was cut with NotI, size-selected to >25 kb, and cloned into the pJAZZ vector. The size of uncut clones was 25-35 kb for *L. helveticus* and 40-60 kb for *E. coli*, corresponding to inserts of 10-20 kb or 25-45 kb, respectively.

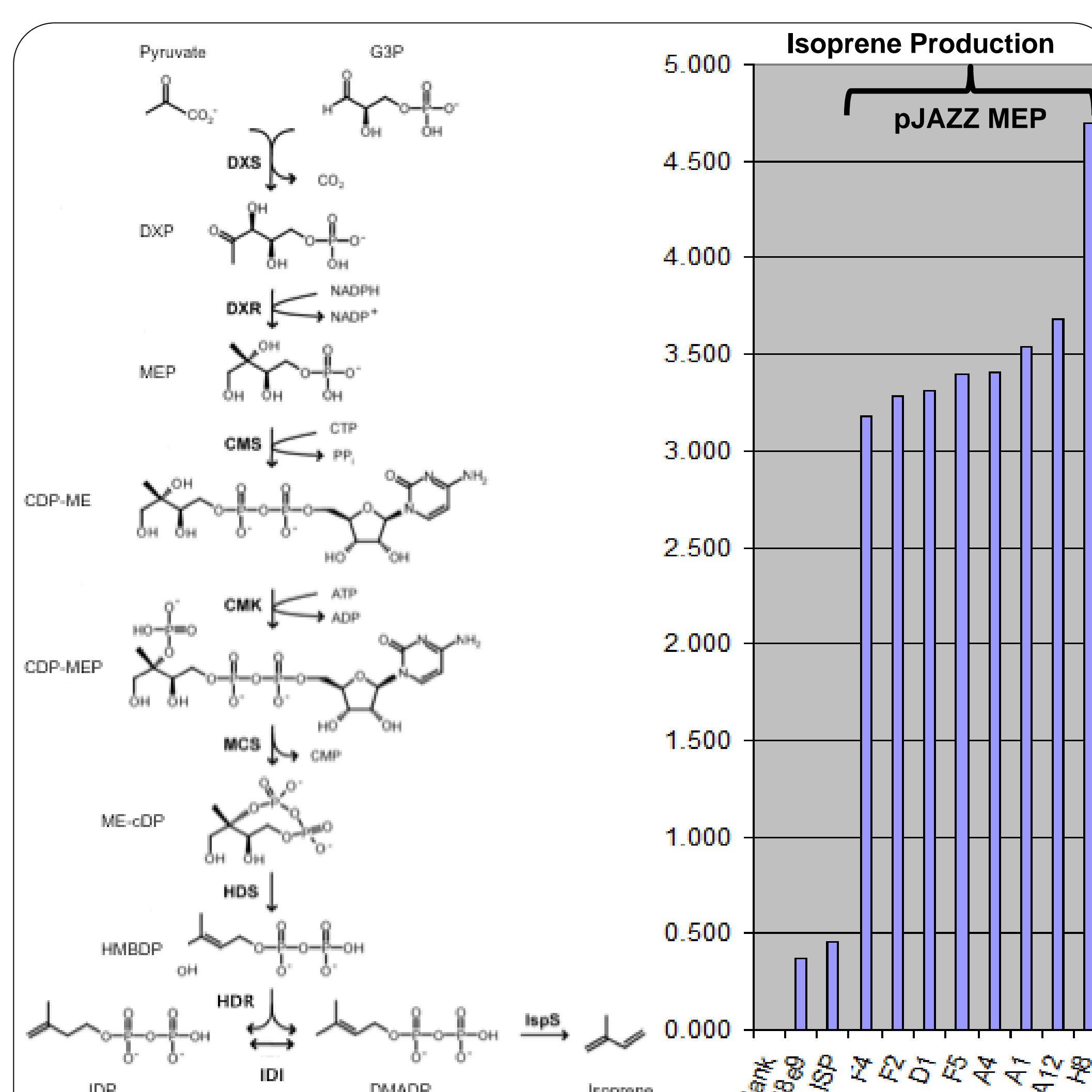
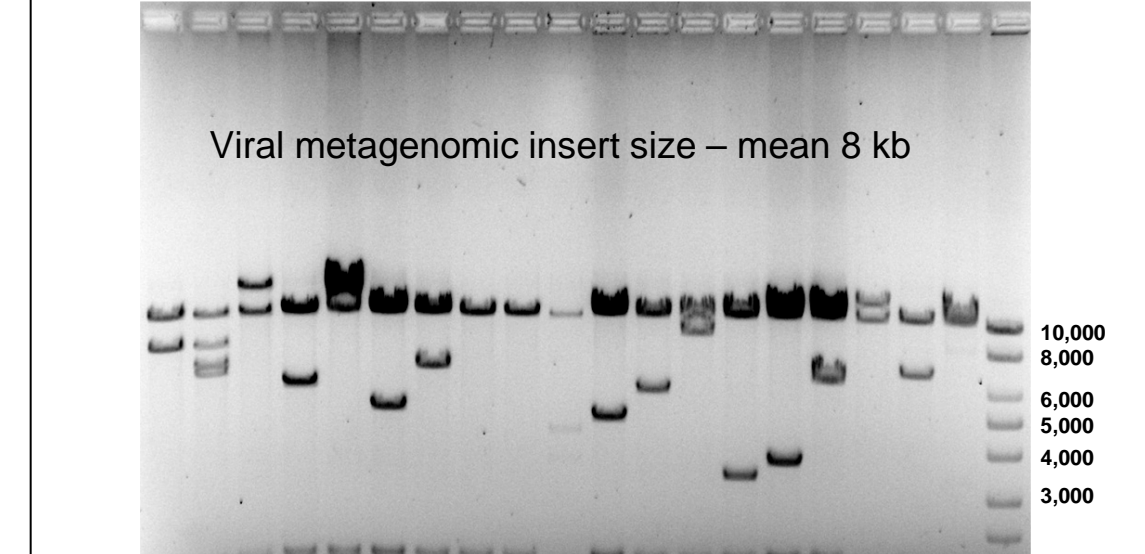


Figure 6. The MEP pathway (left) showing conversion of DMADP into isoprene by isoprene synthase. The MEP pathway encodes 9 genes and results in the product of isoprene, a valuable feedstock chemical that is capable of playing a central role in the future bio-economy. Efforts to clone a 9 kb synthetic MEP pathway were unsuccessful except in the linear vector pJAZZ behind a rhamnose promoter. A 96-well plate containing these clones was screened by University of Wisconsin Stevens Point scientists (Dr. Eric Singaas, PI) for isoprene production (right panel). The last eight data points show that the best pJAZZ isoprene producing clones are ~ 10 fold better than previous constructs.

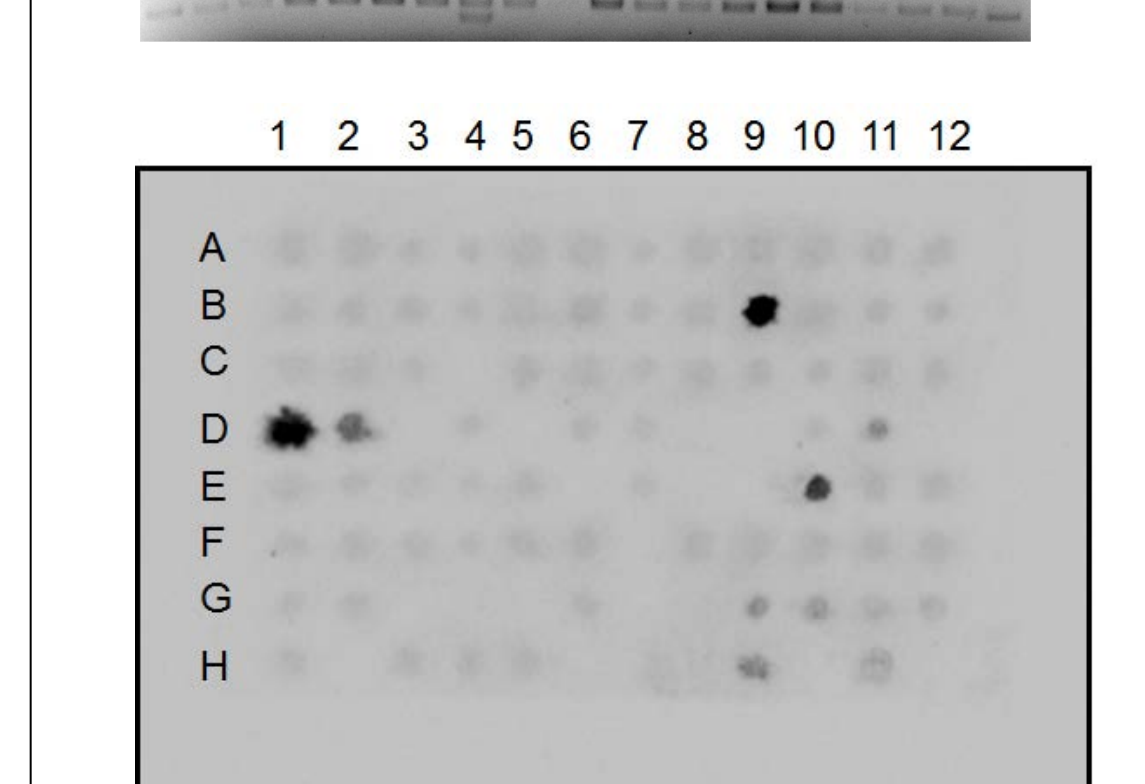
Octopus Hot Spring 93°C



Figure 7. Functional cloning of viral metagenomic replisomes and accessory proteins. Near boiling water from Octopus Hot Spring (left) was filtered (YNP permit # xxxxxx) and the microbial fraction was separated from the viral fraction (Schoenfeld et al. 2008. Assembly of Viral Metagenomes from Yellowstone Hot Springs. Appl. Environ. Microbiol. 74:4164-4174).

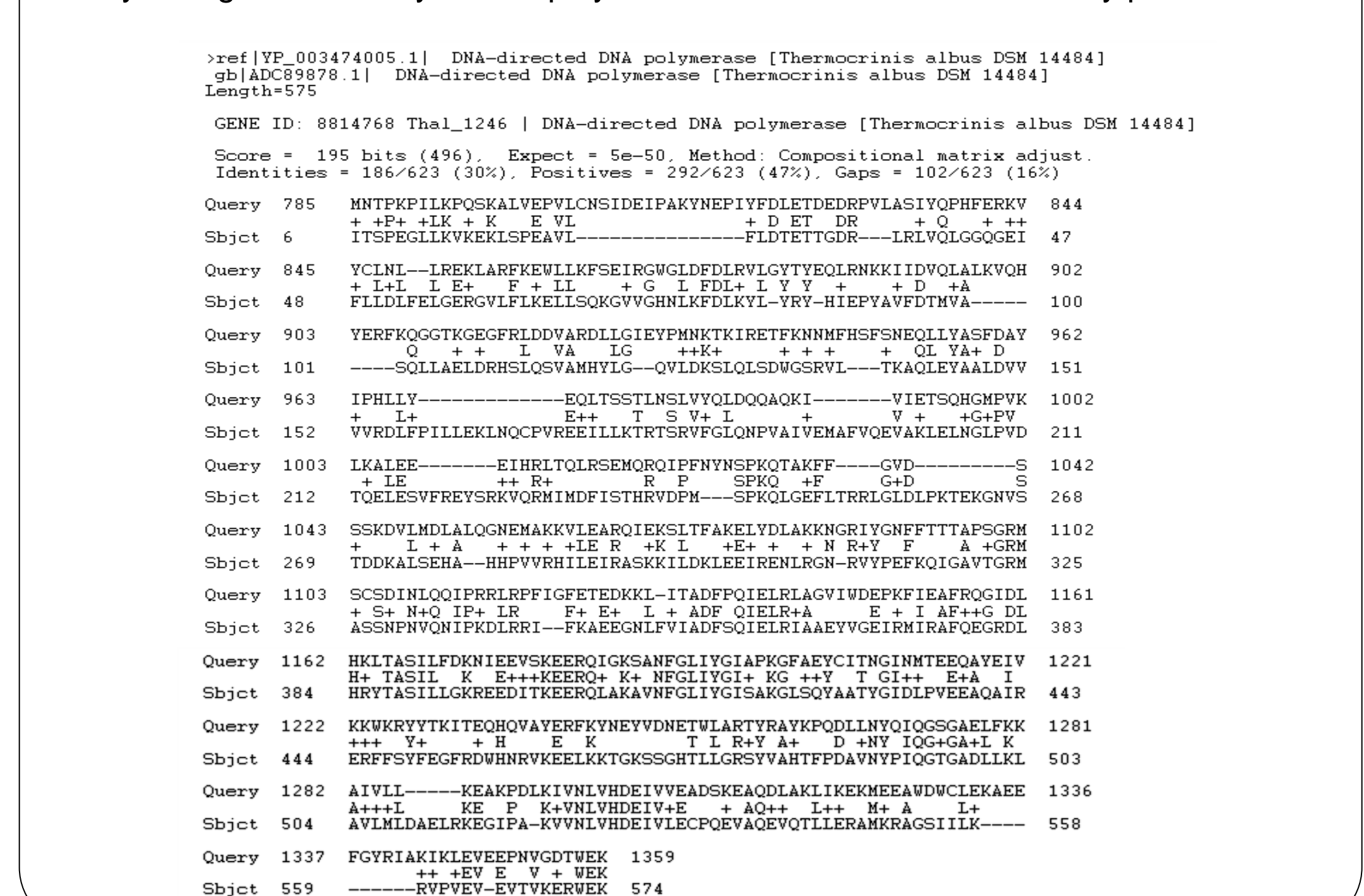


The viral metagenomic DNA was purified, amplified and cloned into the linear pJAZZ vector and transformed into the TSA *E. coli* host strain. Viral metagenomic clones were checked for insert size, which averaged (10.2 kb).



Octopus Hot Spring viral metagenomic clones in pJAZZ were grown overnight, lysed, heat treated, and the presence of thermostable DNA polymerase was assayed by incorporation of isotope in a DNA primer/template reaction mix. Four strong signals were observed.

Two DNA polymerase positive clones were chosen for complete sequence analysis, yielding a new family A DNA polymerase and associated accessory proteins.



CONCLUSION

Two new tools for expression cloning of genes and pathways from genomes and metagenomes has accelerated our ability to ferment platform chemicals such as isoprene and access new enzymes for the breakdown of carbohydrates or the synthesis of nucleic acids. These tools, the Expresso® line of vectors and competent cells, and the pJAZZ® linear cloning system, are available for research use exclusively from Lucigen Corp.