



DNA Synthesis

at the

DOE Joint Genome Institute (JGI)

Ernst Oberortner
eoberortner@lbl.gov

Oct 16th, 2015

Outline

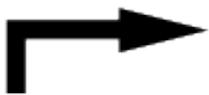
- **DNA Synthesis Program at the DOE JGI**
<http://jgi.doe.gov/our-science/science-programs/synthetic-biology/>
- **Experience Reports**
 - Data Exchange and Standards

Outline

- **DNA Synthesis Program at the DOE JGI**
<http://jgi.doe.gov/our-science/science-programs/synthetic-biology/>
- **Experience Reports**
 - Data Exchange and Standards

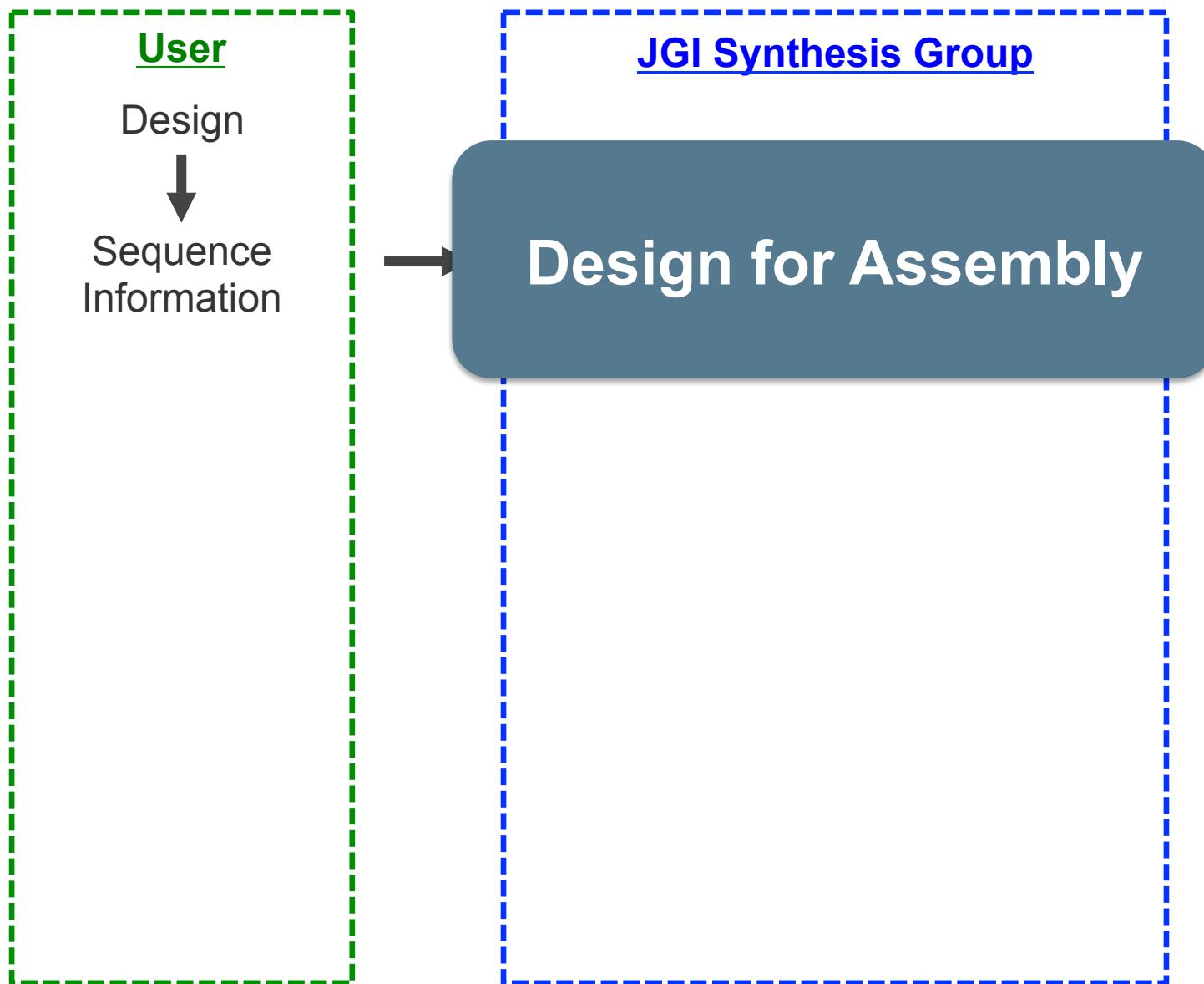
DNA Synthesis Workflow

- Hypothetical Example:



ttgacggcta . . . acagtgc tagc TCTAGAGAAA . . . AAACTAGATG atgagtgtatcgcta . . . cgcaaacctgtggtcgcctaataa ccaggcatc . . . tgtcggtaacgctctc

DNA Synthesis Workflow

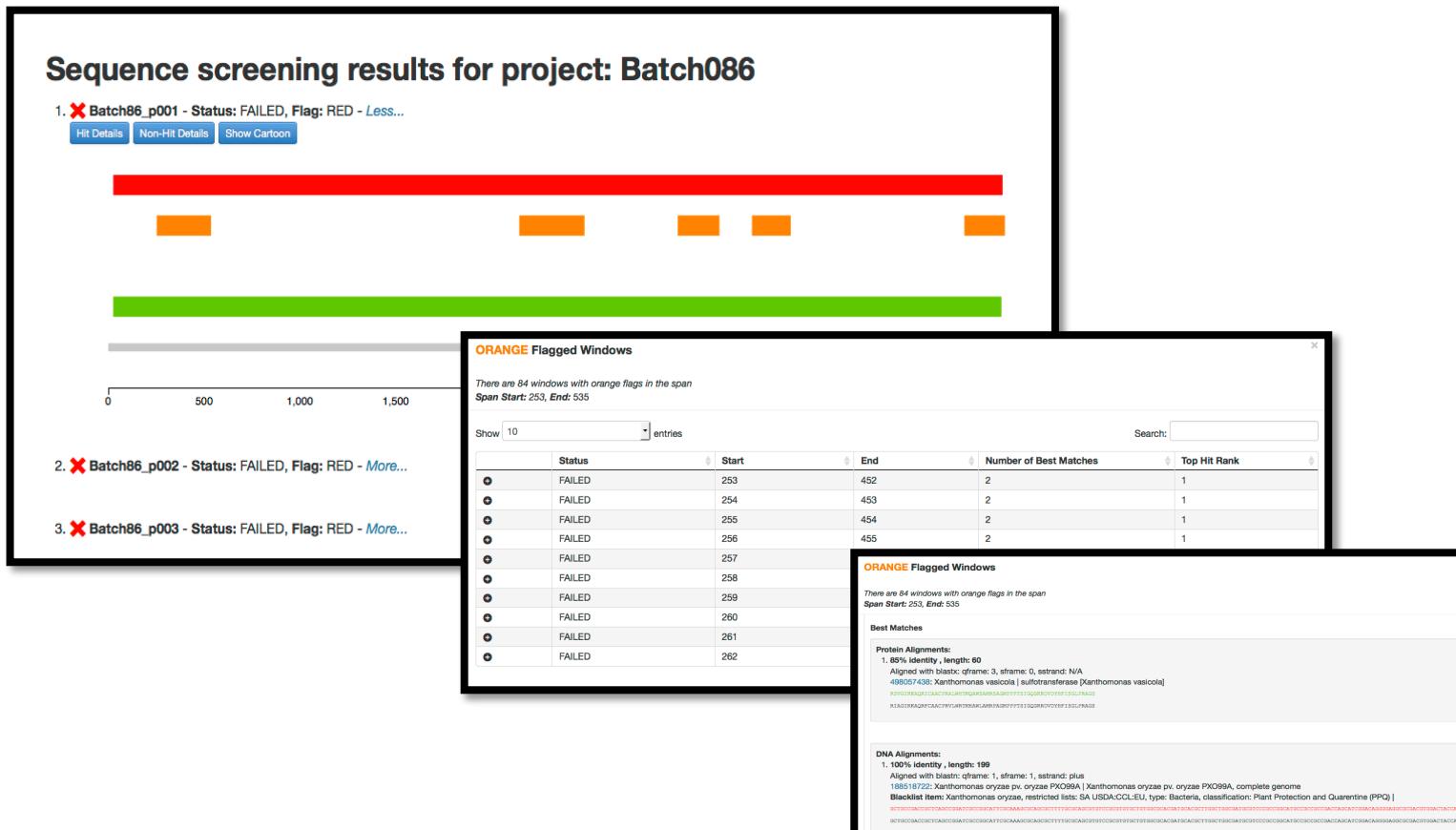




Lisa Simirenko

- **Design Task: Screen Sequences**

- BLISS detects “sequences of concern” of at least 200 nucleotides in length on either DNA strand, including polypeptide translations using the three alternative reading frames on each DNA strand (six-frame translation).



- Design Task: “Polish” Sequences against DNA Synthesis Constraints

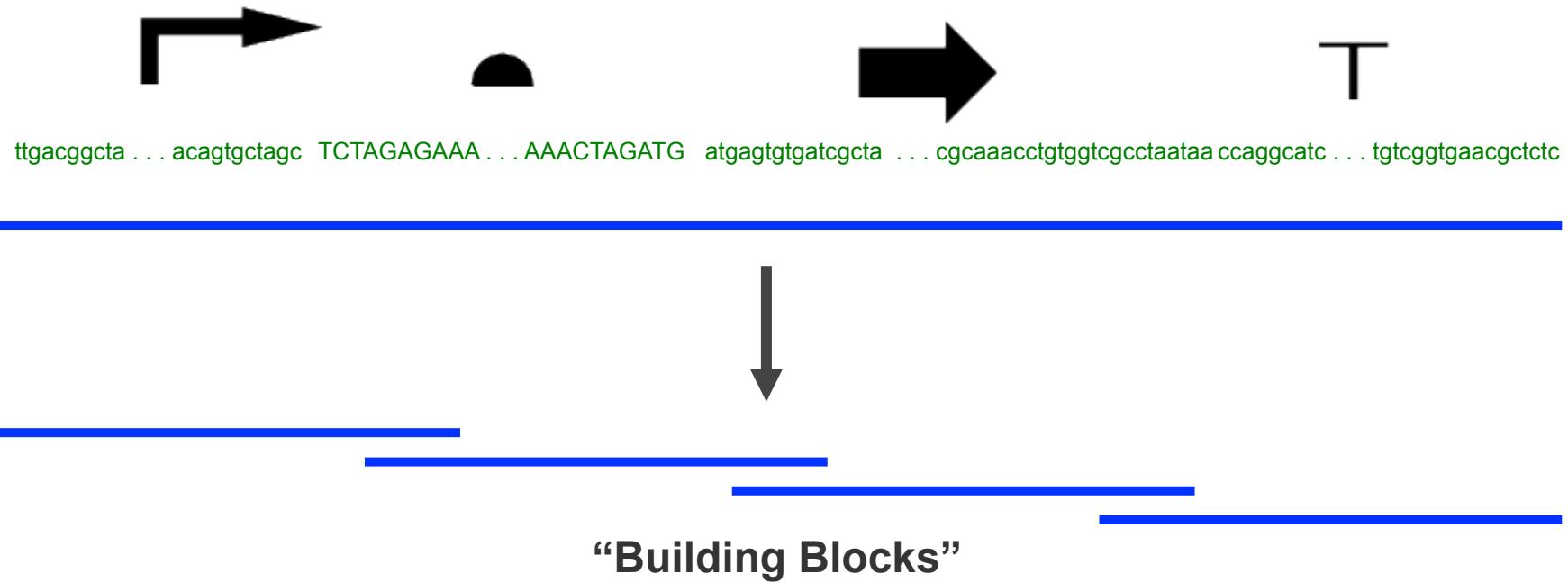


ttgacggcta . . . acaggtgctagc TCTAGAGAAA . . . AAACTAGATG atgagtgtgatcgcta . . . cgcaacctgtggtcgcctaataa ccaggcatc . . . tgtcggtaacgcctc

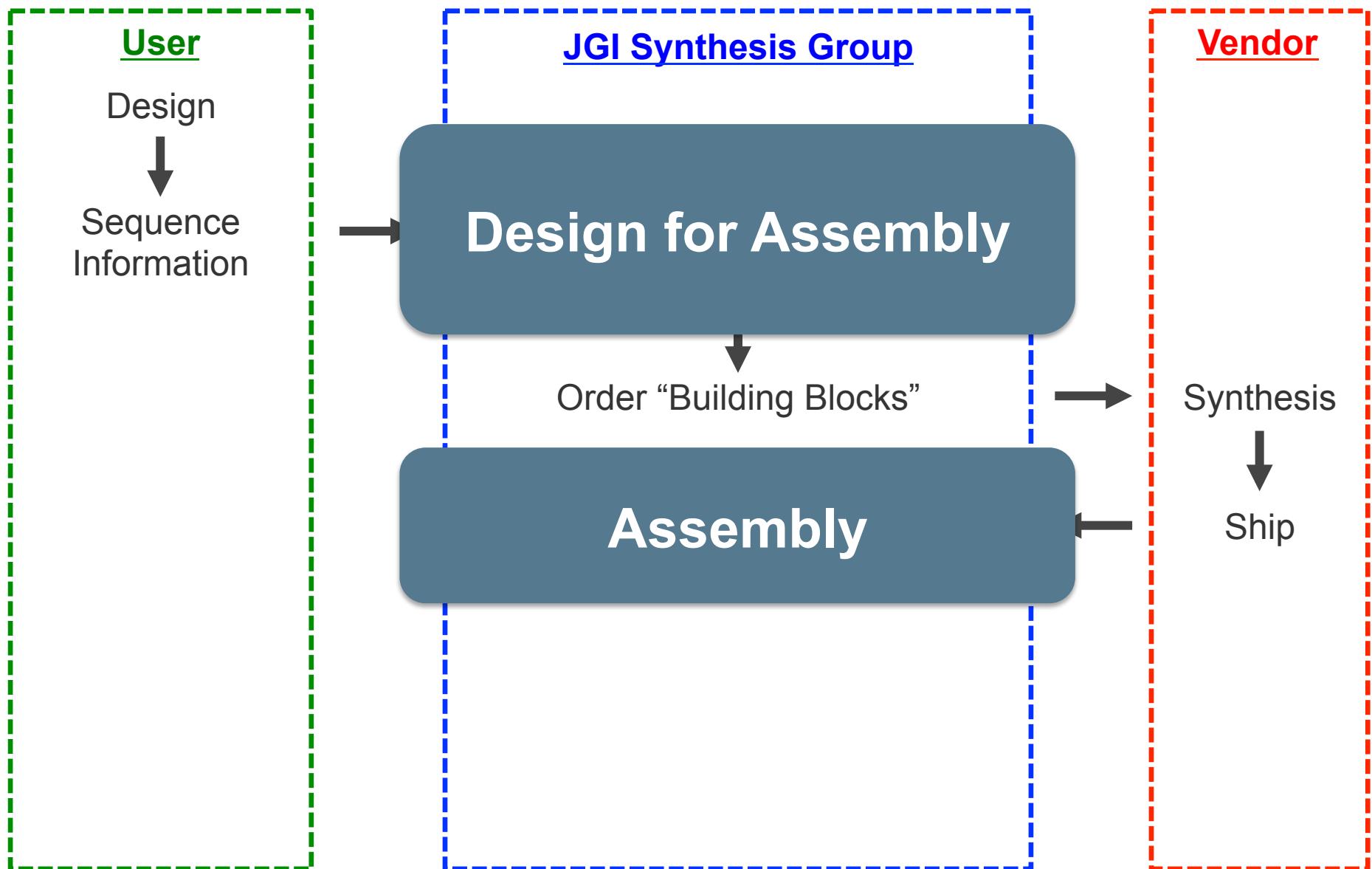
- **GC% Content**
 - Global vs. Local (“Windowed”)
 - GC% Flux
- **Repeats**
 - Global vs. Local (“Windowed”)
 - Tandem vs. Interspersed
 - Direct vs. Inverted
 - Exact vs. Mutated
- **Repeat Coverage**
 - regions with high repeat content
- **Restriction Sites**

Design for Assembly

- Design Task: Decompose Sequences



DNA Synthesis Workflow

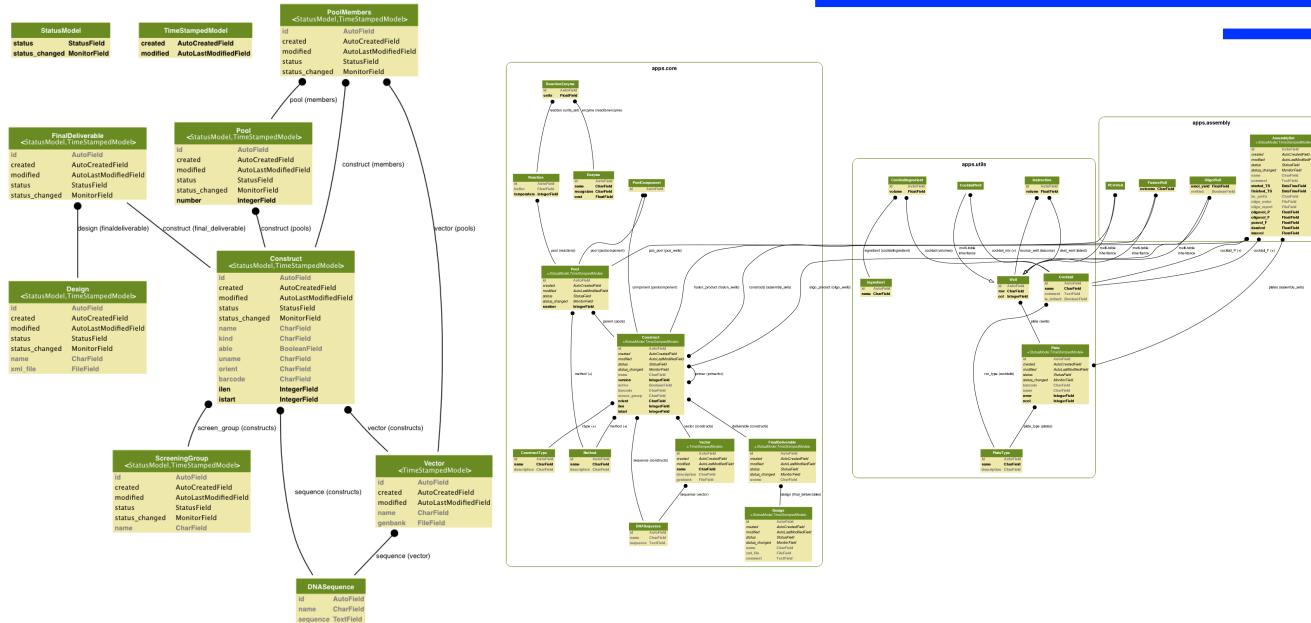


Tracking of the Assembly Process



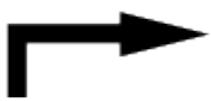
John Meng

“Building Blocks”

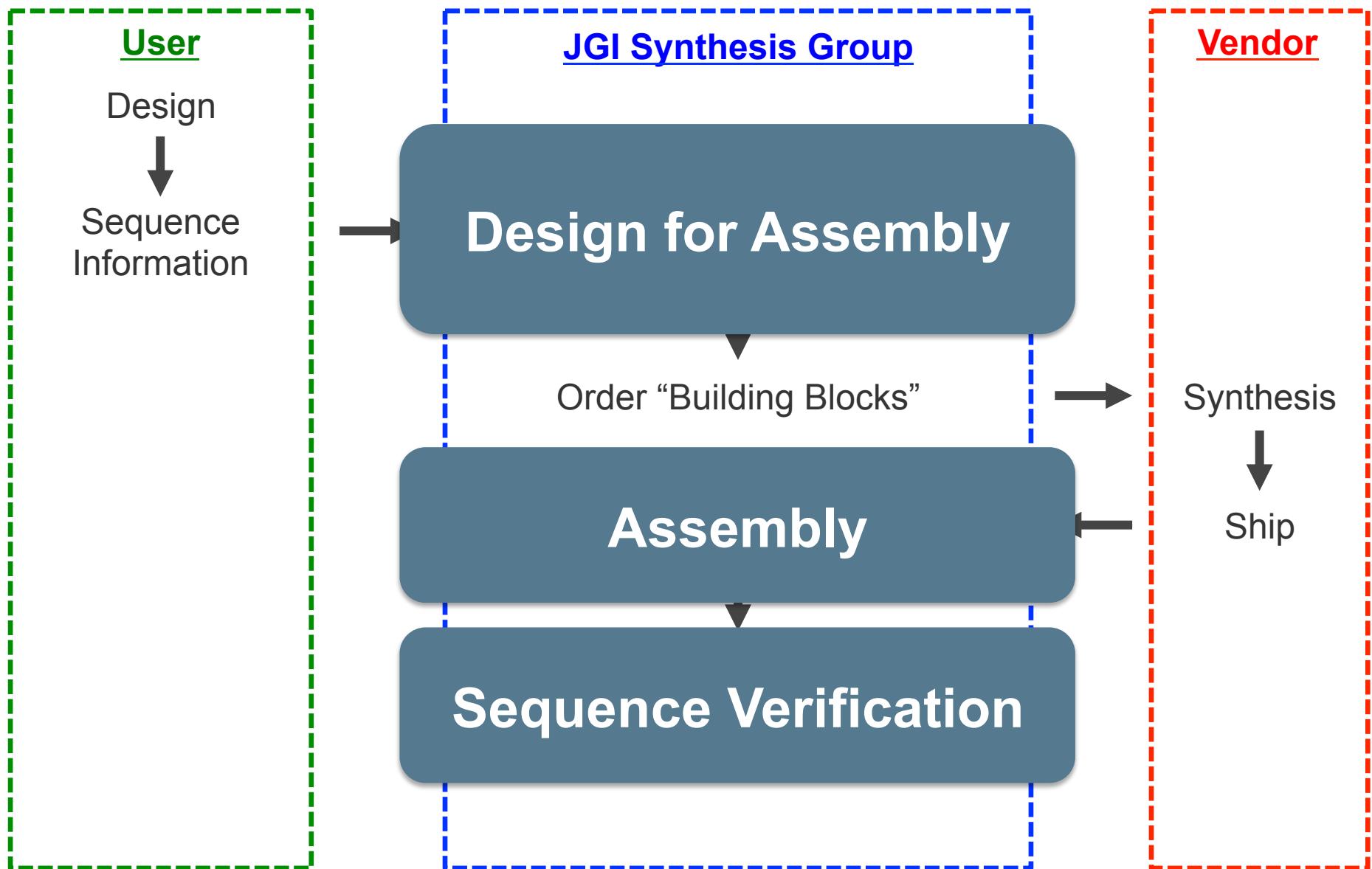


SynTrack

ttgacggcta . . . acagtgttagc TCTAGAGAAA . . . AAACTAGATG atgagtgtatcgcta . . . cgcaaacctgtggtcgcctaataa ccaggcatc . . . tgtcggtgaacgcctc



DNA Synthesis Workflow



Current Sequencing Technologies

PacBio



Illumina

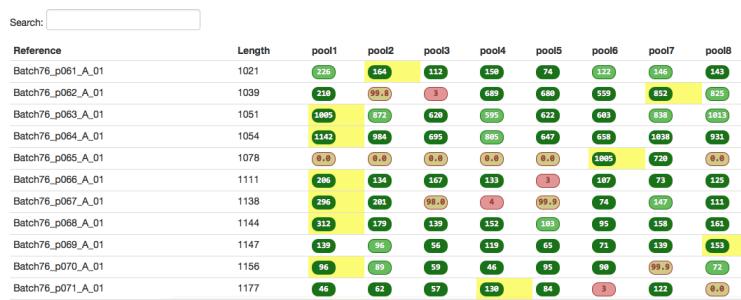


Pipeline

Analysis: P100814

Pool	SampleID	roi	sub	IGV links	Excel file	Call
pool1	HAR35893	roi	sub	igv.xml		
pool2	HAR35894	roi	sub	igv.xml		
pool3	HAR35895	roi	sub	igv.xml		
pool4	HAR35896	roi	sub	igv.xml		
pool5	HAR35897	roi	sub	igv.xml		
pool6	HAR35898	roi	sub	igv.xml		
pool7	HAR35899	roi	sub	igv.xml		
pool8	HAR35900	roi	sub	igv.xml		

Results



Representation of Sequencing Results

Analysis: AAONP

Overview

Pool	SampleID	bwa_dir
1480_1066562_AAHPB	1480_1066562_AAHPB	bwa_dir
1480_1066562_AAHBS	1480_1066562_AAHBS	bwa_dir
1480_1066562_AAHBO	1480_1066562_AAHBO	bwa_dir
1480_1066562_AAHBH	1480_1066562_AAHBH	bwa_dir
1480_1066562_AAHBN	1480_1066562_AAHBN	bwa_dir
1480_1066562_AAHBG	1480_1066562_AAHBG	bwa_dir
1480_1066562_AAHBH	1480_1066562_AAHBH	bwa_dir
1480_1066562_AAHBC	1480_1066562_AAHBC	bwa_dir

IGV links	
	igv.xml

Excel Sheet

Excel file	
Perfect	101 (45.5%)
Fixes	14 (6.3%)
Failed	107 (48.2%)

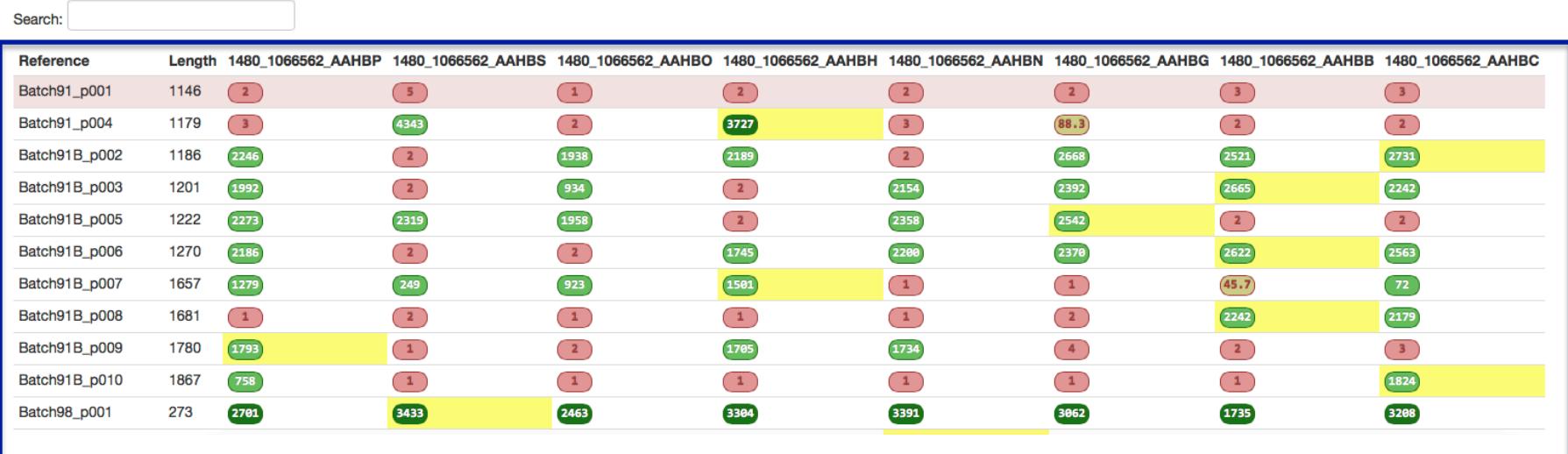
Statistics

Legend

Call		
F	Flawless	everything is looking good
A	Almost	all variants within 10 bases
I	Incomplete	not all positions covered
L	Low coverage	mean coverage < 30
E	Errors	variants
D	Dips	Dips
?	No call	

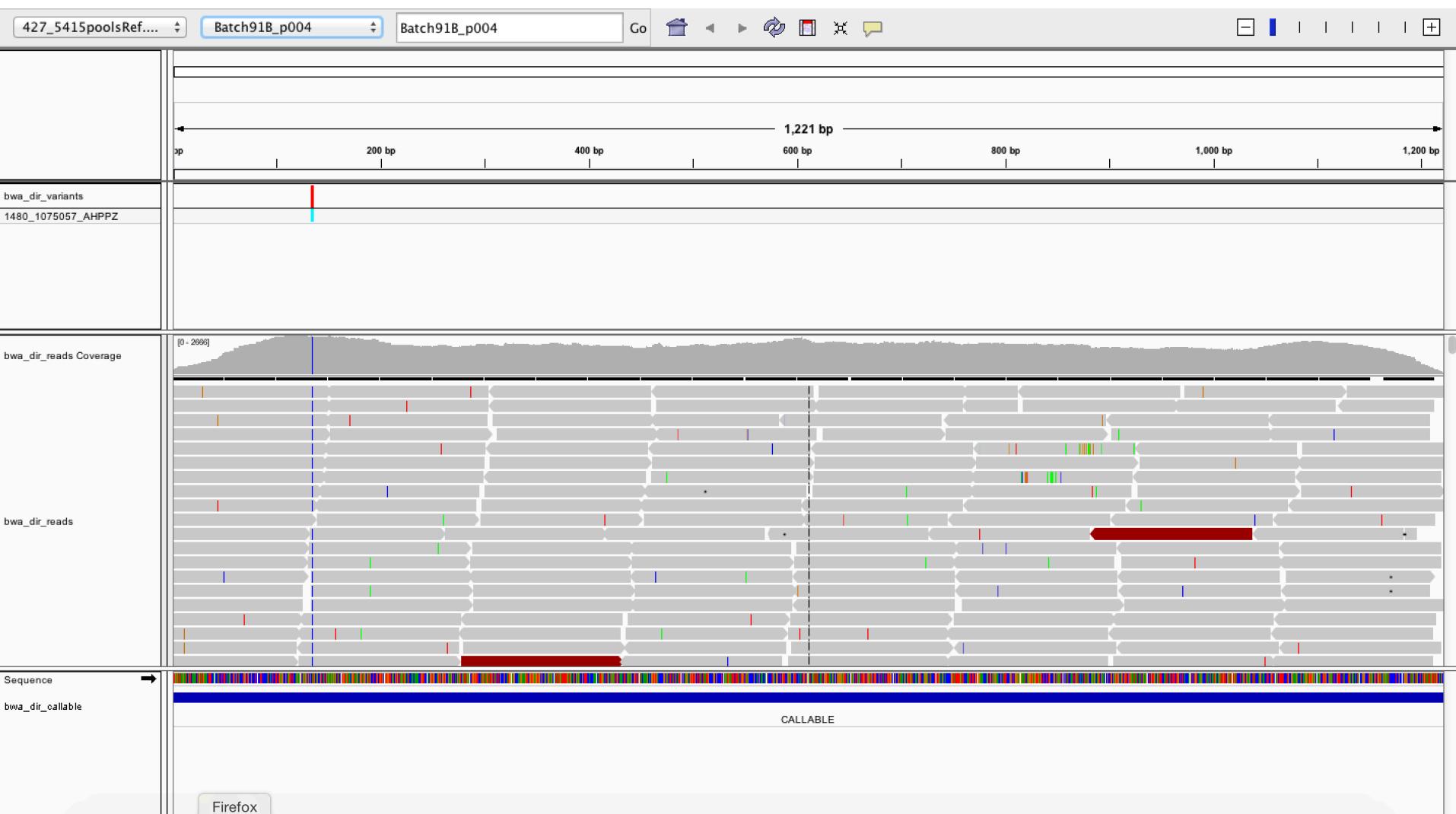
Results

IGV

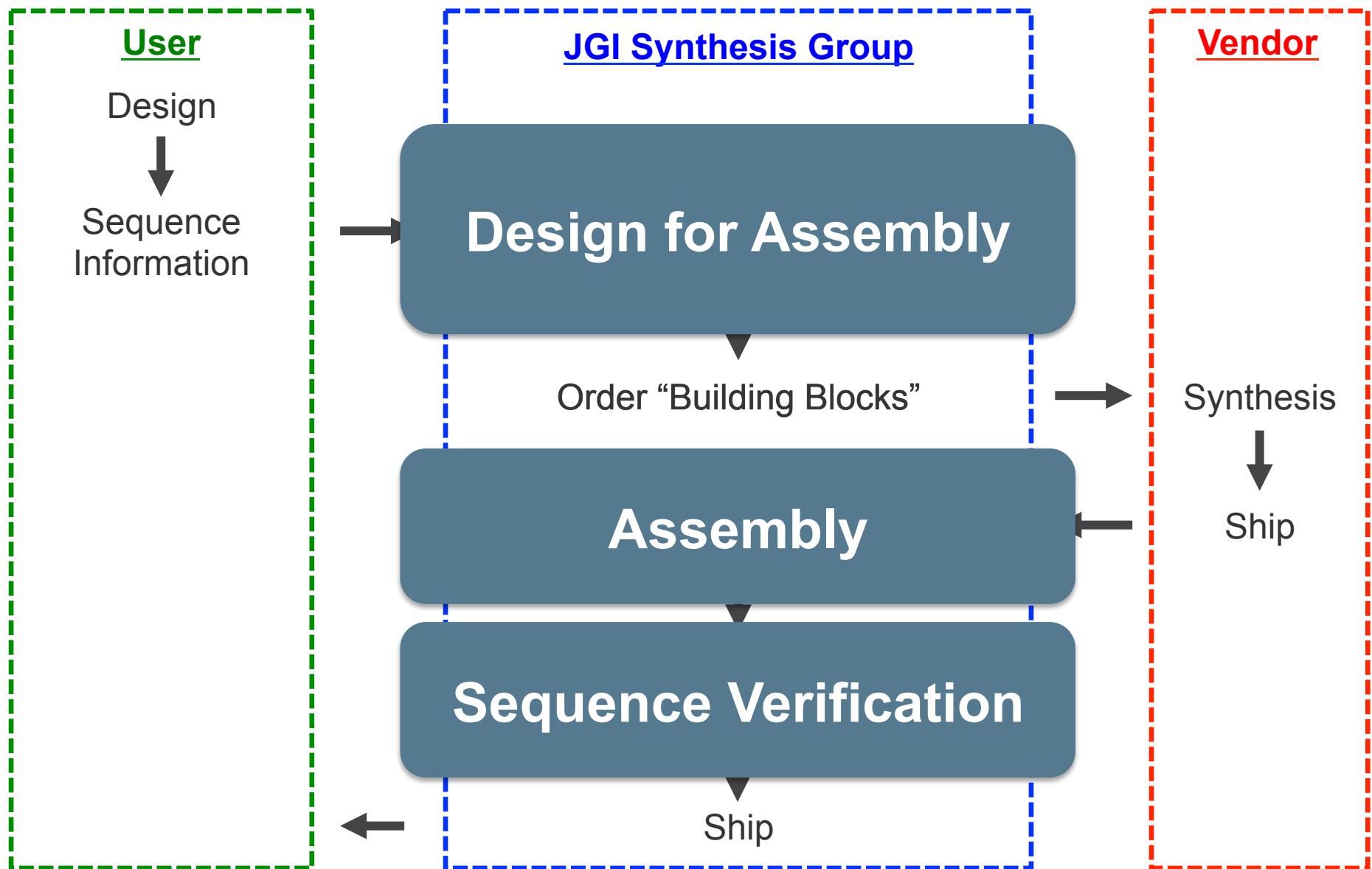


Sequencing Results

Integrated Genome Viewer (IGV)

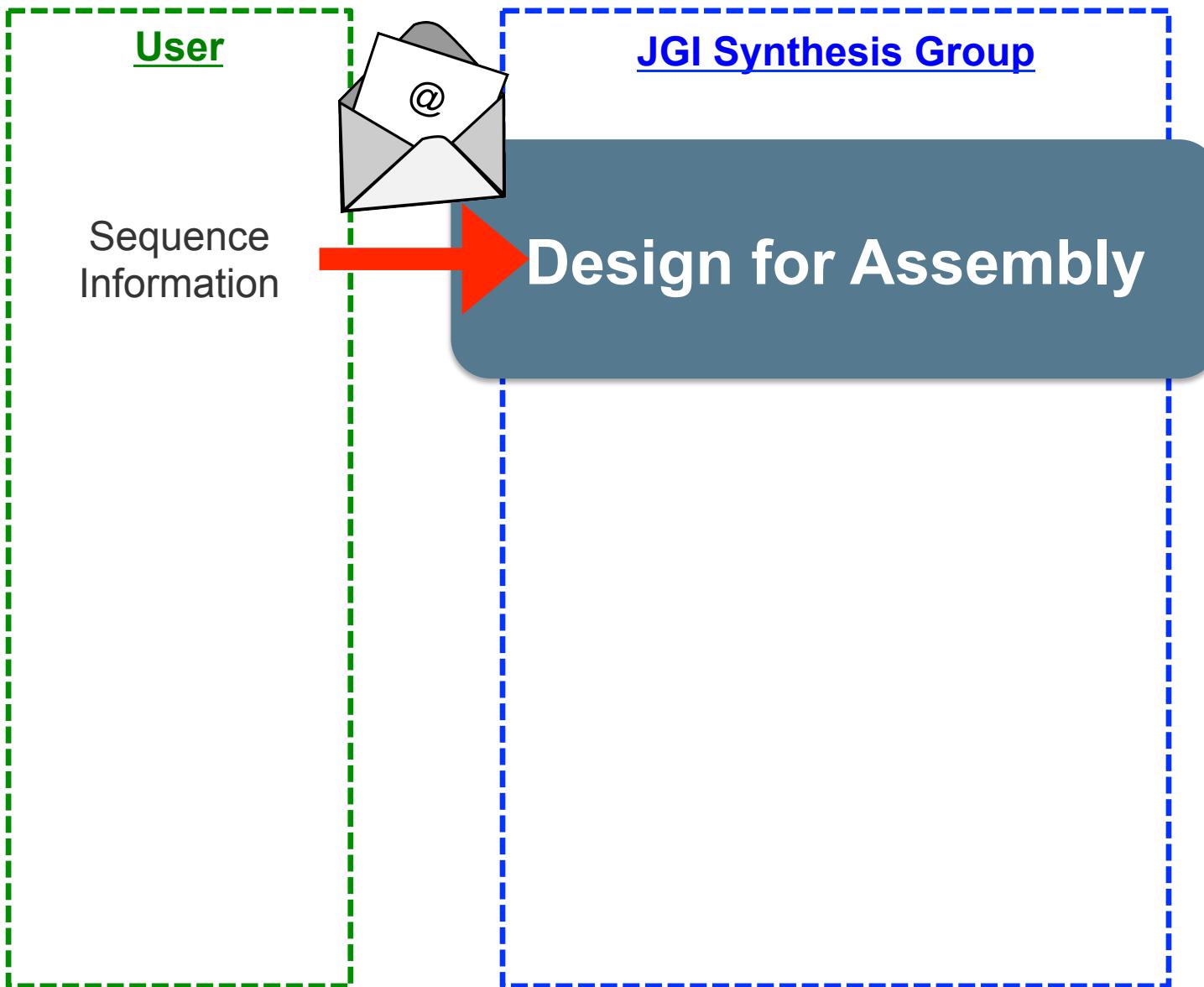


DNA Synthesis Workflow



- **DNA Synthesis Program at the DOE JGI**
<http://jgi.doe.gov/our-science/science-programs/synthetic-biology/>
- **Experience Reports**
 - Data Exchange and Standards

Current: Sequence Submission



- “Outliers”

- Genbank-like format as Word Document including color-coded sequence annotations

```

LOCUS
DEFINITION
FEATURES      Location/Qualifiers
aa            1..2
              /name="aa-name"
bb            3..4
              /name="bb-name"
cc            5..6
              /name="cc-name"
misc          1..10
              /misc="restriction site"
  
```

ORIGIN



- “Outliers”
 - FASTA-like format including natural language descriptions of composition

Set 2 expression of 6- **XX genes** in **YY** using constitutive like pBAD / or inducible expression vector **ZZ** replicative plasmid **will harbor** six genes in the format:
Cons. P-R-G1, R-G2, R-G3, R-G4, R-G5, R-G6-T,

G1

ATG . . . TAA

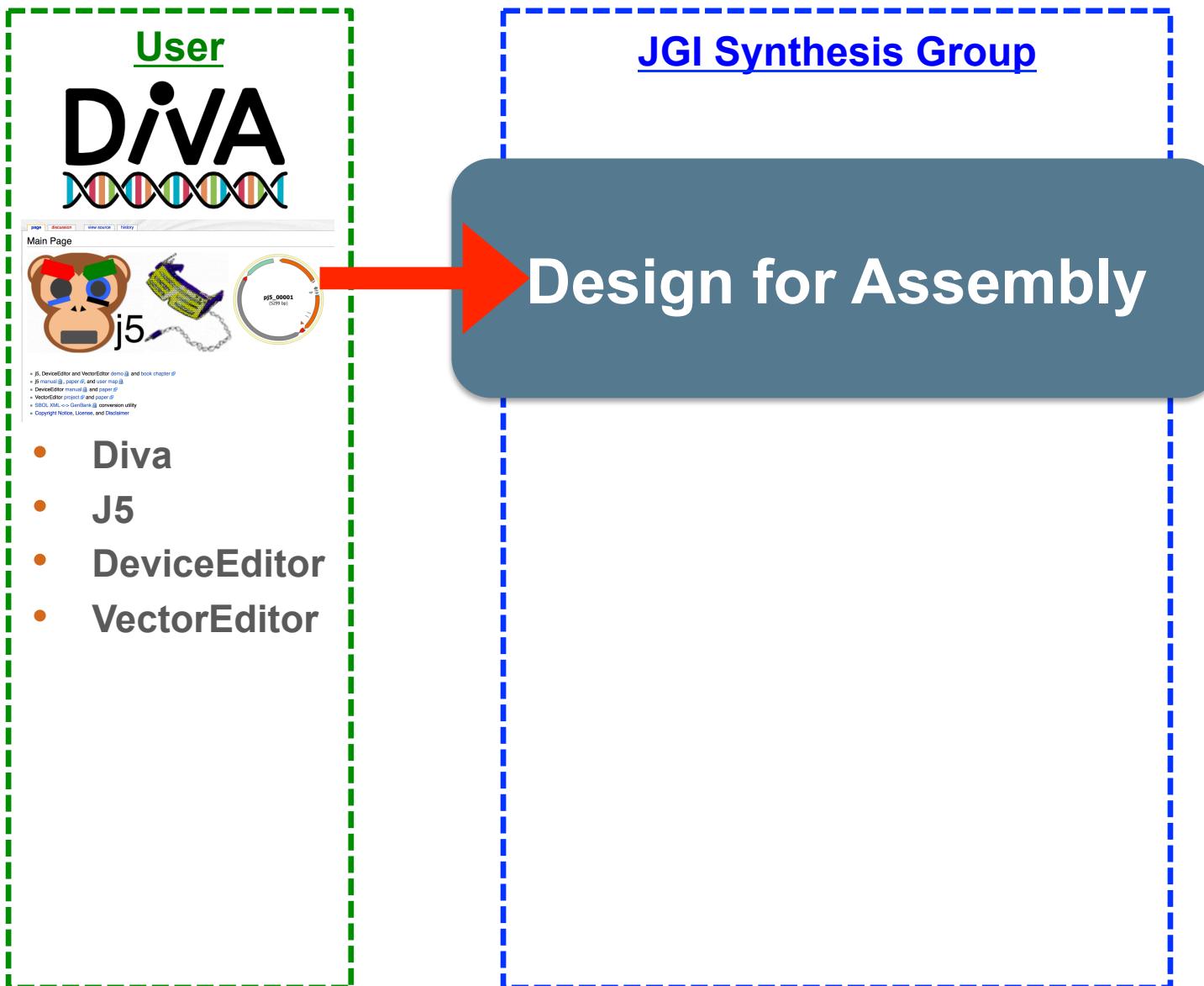
G2

ATG . . . TAA

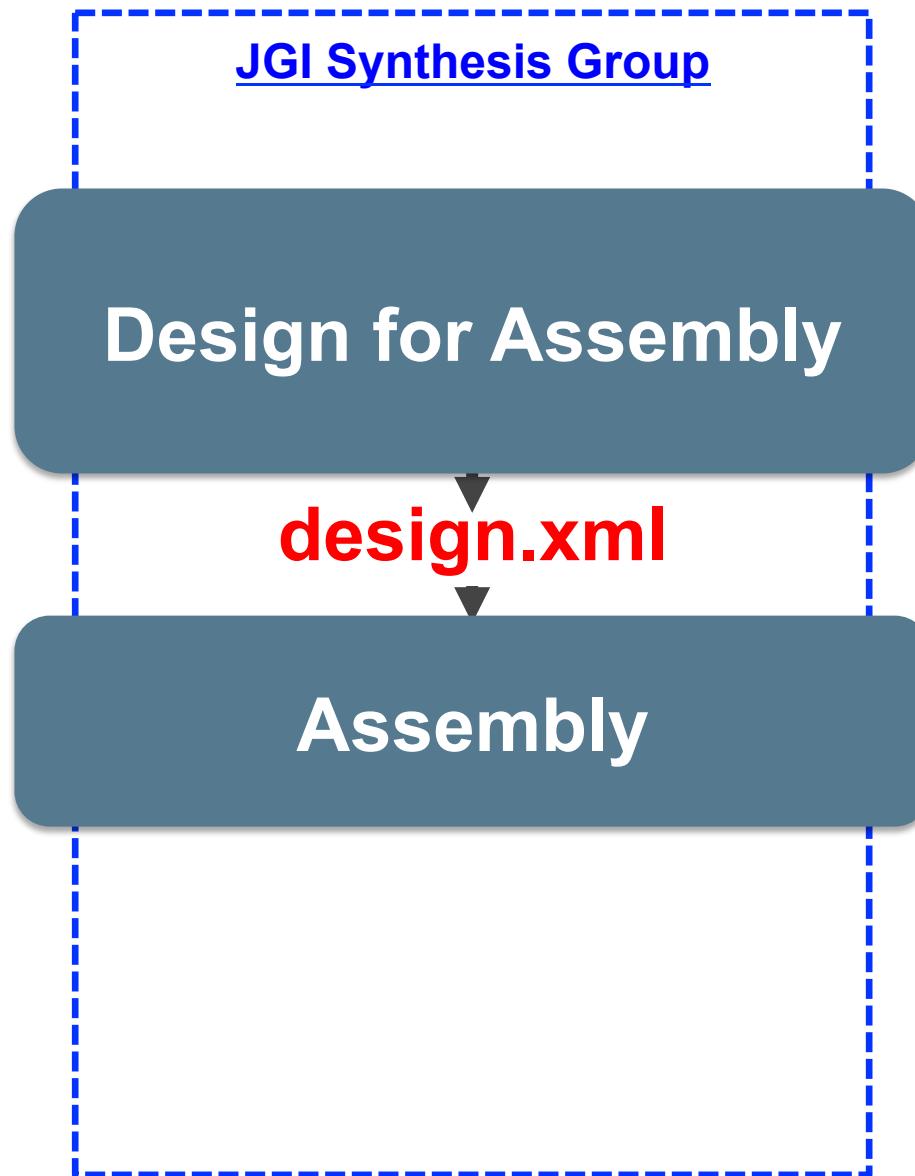
G3

ATG . . . TAA

Vision: Design and Sequence Submission



Adoption of SBOL



DOE JGI's design.xml

- Supported by SBOL v1.1.0

<design>

<constructs>

Collection

<construct>

DNAComponent

<subconstructs>

SequenceAnnotation

<pool>

<subconstruct>sub1</subconstruct>

DNAComponent

<subconstruct>sub2</subconstruct>

DNAComponent

</pool>

</subconstructs>

<sequence></sequence>

DNASequence

</construct>

<construct id="sub1">
 <sequence></sequence>
</construct>

<construct id="sub2">
 <sequence></sequence>
</construct>

DOE JGI's design.xml

- Not supported by SBOL v1.1.0

```
<design id="...">
  <constructs>
    <construct kind="..." id="..." method="..." deliverable="...">
      <uid>UID</uid>
      <sequence>ATGTA</sequence>
      <GCp>0.2</GCp>
      <seqlen>5</seqlen>
      <vector>vector</vector>
      <subconstructs>
        <pool number="..." method="...">
          <reaction temperature="..." buffer="...">
            <enzyme name="..." units="..." />
          </reaction>
          <subconstruct> </subconstruct> Assembly-specific
        </pool> Information
      </subconstructs>
    </construct>
```

Adoption of SBOL – Vision



- **Adopt SBOL v2.0.0**
 - Computed Values: Annotations (Core)
- **Contribute to SBOL v2.0.0**
 - Development and Evolution
 - Assembly-specific Information
 - Sequence Verification / QC
- **Use, Development, and Evaluation of libSBOLj v2.0**

Acknowledgements

