TelesisBio

BioXp® Error-Corrected Library Synthesis Solution User Guide

Part number 43058 | Revision 1.2 | Effective 29NOV2022

Contents

Introduction to error corrected library synthesis on the BioXp® system

The BioXp system enables the synthesis of error-corrected libraries (ECLs). ECLs can be synthesized up to a diversity of 64 million variants per well (assuming all 20 amino acids across six different degenerate codon positions such as NNK) using the BioXp system. For this workflow, digital sequences between 300bp and 800bp are the inputs and ECLs are the output of an overnight run on the BioXp system. Sequences between 800bp and 2kb are also orderable through myBioXperience (see section on complexity and sequence submission requirements below). The ECL synthesis kit contains all the necessary components to produce DNA libraries through proprietary processes that involve *de novo* DNA synthesis, error correction and amplification. The final product is the purified ECLs in TE buffer (10mM Tris, 1mM EDTA pH8) that are ready for downstream applications such as cloning and screening or selection.

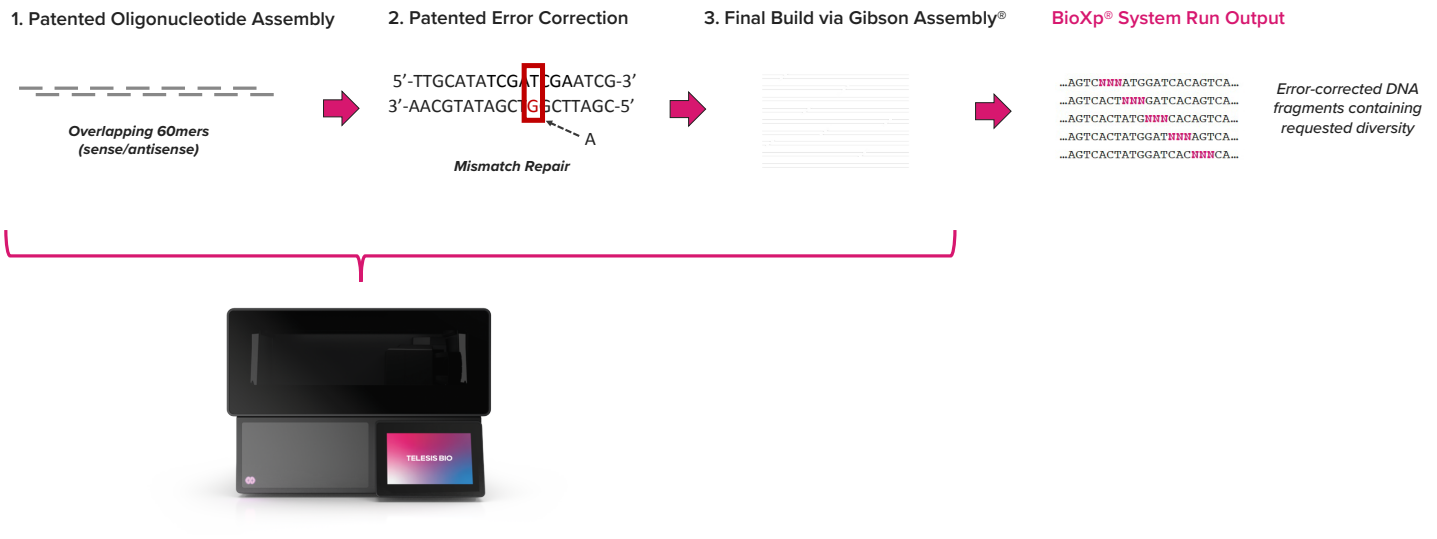


Figure 1. Graphical representation of the end-to-end steps to generate an error-corrected library from a digital sequence on the BioXp 3250 system.

Overview of the BioXp Error corrected library synthesis workflow

Types of libraries

The BioXp system enables the synthesis of a variety of ECLs. Specifically, both scanning and combinatorial libraries can be synthesized on the BioXp system. Scanning ECLs can be further categorized into simple-scanning (single amino acid/codon substitutions per well) and site-saturation error corrected libraries that employ degenerate codons such as (NNN, NNK and NNS). Combinatorial ECLs can include additional degenerate bases from the IUPAC code. The number of degenerate bases, codons and their locations within the library are subject to the ECL product specifications described below in the Input Requirements section.

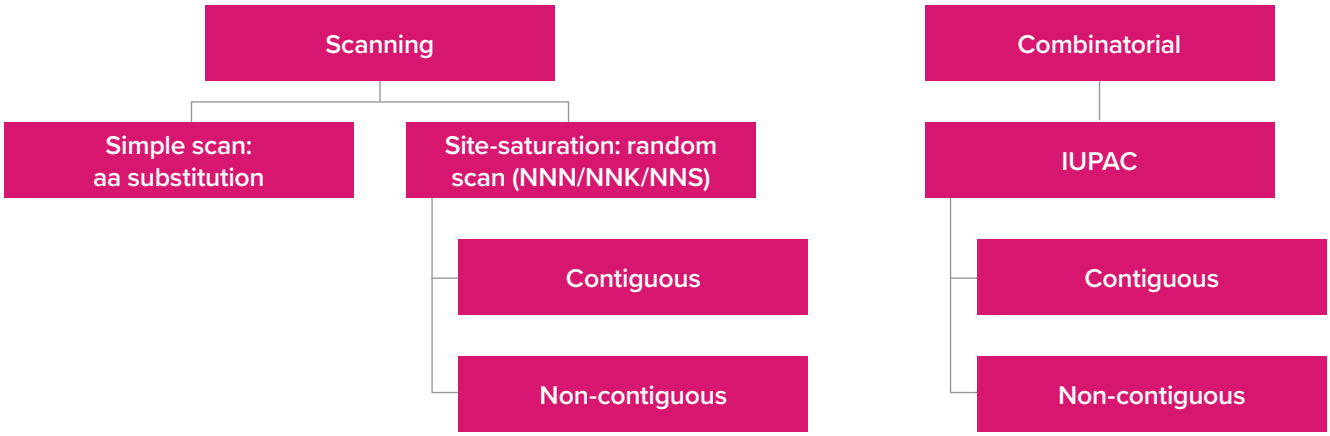


Figure 2: Schematic representation of the types of variant libraries supported on the BioXp system.

Input requirements and guidelines for submission of DNA sequences

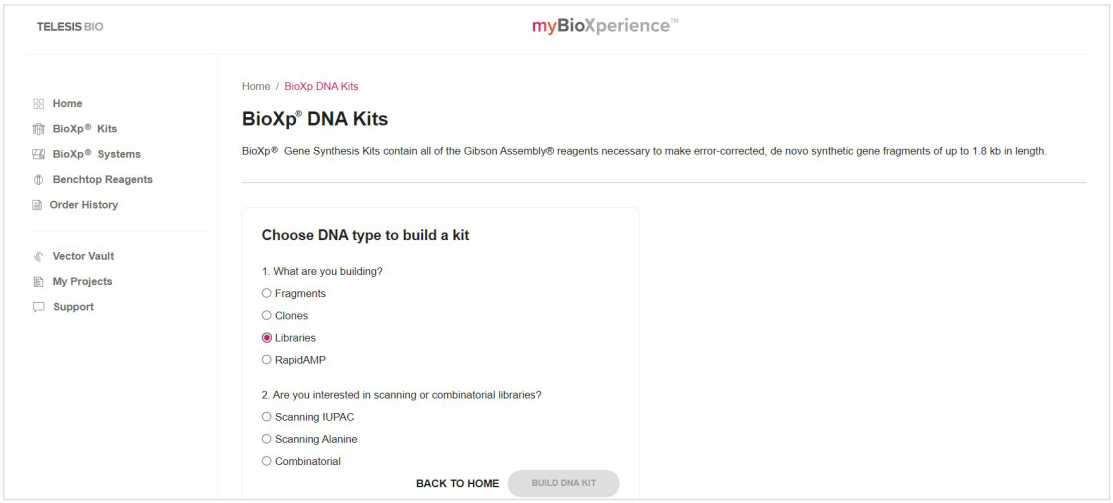
During the ordering process, the input sequence will be subject to complexity check in the portal to inform customers with the "buildability" information. Each sequence will be subject to specific requirements to enable accurate prediction about synthesis on the BioXp system. Each analyzed sequence will be ascribed a color ("Green", "Yellow" or "Black") as a result of this analysis. Detailed description of sequence requirements are listed below and the definitions of color assignments are listed in the "Sequence Submission" section.

1. Acceptable sequence length: Sequence length 300 – 800bp will be scored "Green" in complexity check and 801 – 2000bp will score "Yellow" since the latter is outside of the current complexity requirements but allowed to proceed for ordering. Sequences >2000bp will be flagged as "Black" and not allowed for ordering.
2. Acceptable sequence GC %: 20% – 70%. Sequence GC% between 30% – 60% will be "Green" and the GC% between >20% to <30% and >60% to <70% will be "Yellow" since the latter is outside of the current complexity requirements but allowed to proceed for ordering. GC% outside of this range will be flagged as "Black" and not allowed for ordering.
3. In terms of sequence identity, the variants should be within 85% when compared against the reference sequence. Sequence identity >85% will be "Green" and allowed to proceed for ordering. Sequence identity <85% will be flagged as "Black" and would need redesign.
4. The degenerate codons or bases cannot be present within the first or the last 60 bases of the variant sequences. When the degenerate bases are present outside the first and last 60 bases from the reference sequence, the sequences will be "Green" and allowed to proceed for ordering. If the degenerate bases are present within the first or last 60 bases of the input sequences, they will be flagged as "Black" and would need to be redesigned.
5. For the introduction of degenerate bases, "Green" parameters are listed below. Outside these recommendations, sequences will be flagged as "Yellow" but allowed to proceed for ordering.
    a. For libraries containing degenerate codons:
        i. Three types of degenerate codons are allowed: NNN, NNK and NNS
        ii. Up to three contiguous degenerate codons are allowed per sequence
        iii. Up to six non-contiguous degenerate codons are allowed per sequence
    b. For the combinatorial libraries:
        i. All types of degenerate IUPAC bases are allowed
        ii. Up to three contiguous degenerate IUPAC bases are allowed per sequence
        iii. Up to six non-contiguous degenerate IUPAC bases are allowed per sequence
6. In addition, the complexity parameters are computed by estimating homopolymer and repeat lengths and identities and GC flux within the sequence. These parameters will contribute towards categorizing an input ECL sequence as a "Green," "Yellow" or "Black."

Initiating your Libraries Project and ordering through myBioXperience

Visit the Telesis Bio ordering platform, myBioXperience to initiate a new project

A



B

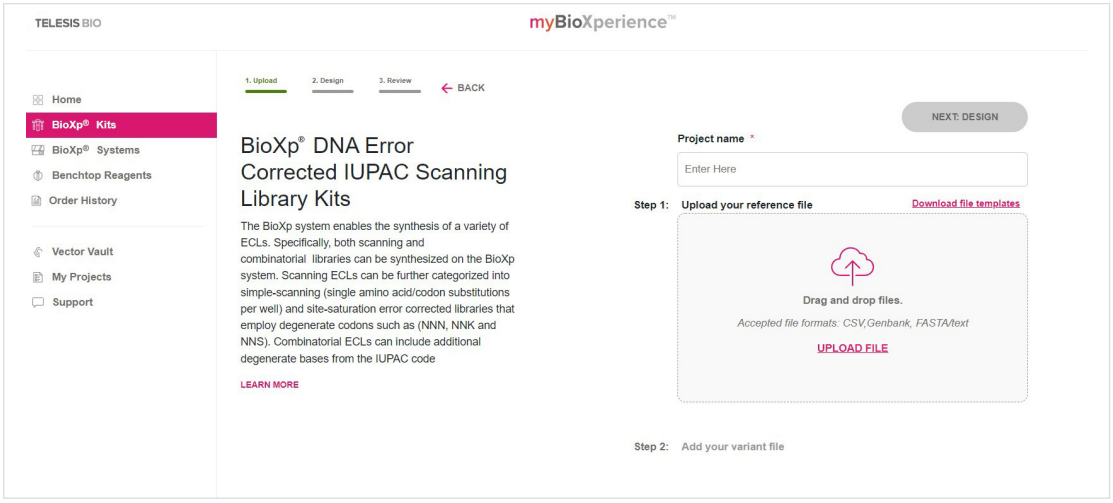


Figure 3: Screenshots of myBioXperience showing A) selection of the ECL synthesis kit options once you have logged into myBioXperience B) selection of the specific type of ECL synthesis kit and upload of the sequences of interest.

Sequence submission and complexity check:

- ☐ Select the DNA application option to get started.
- ☐ Upload your sequence(s) by using the upload tool (Figure 3B). The supported sequence formats are: CSV, Genbank and FASTA/text. Once uploaded, your sequence will be displayed as illustrated in Figure 3.
- ☐ After entering a Project Name, proceed with the complexity check by selecting the button on the right side of the uploaded sequence [show example of button].
- ☐ The complexity check will then analyze your uploaded sequences to ensure that your sequences meet the requirements described below in the **Complexity requirements** section:

  Green = sequences meet specifications, you may proceed to order.

  Yellow = sequences do not meet specifications, you may proceed to order; however, by doing so you accept the risk that the synthesis is not guaranteed as the sequences flagged “yellow” are considered outside of our complexity requirements.

  Black = sequences cannot be ordered.

  Note: If you have an incorrect sequence, then you will get a message that says your sequence is not qualified.

- ☐ Click next to review a summary of your order. Once satisfied you may add the order to the cart.

Understanding the BioXp system

In preparation for initiating the run, it is helpful to understand the various components of the BioXp system deck in order to correctly place all the reagents and consumables into the system. Below is a rendering of the deck with all the key components labeled.

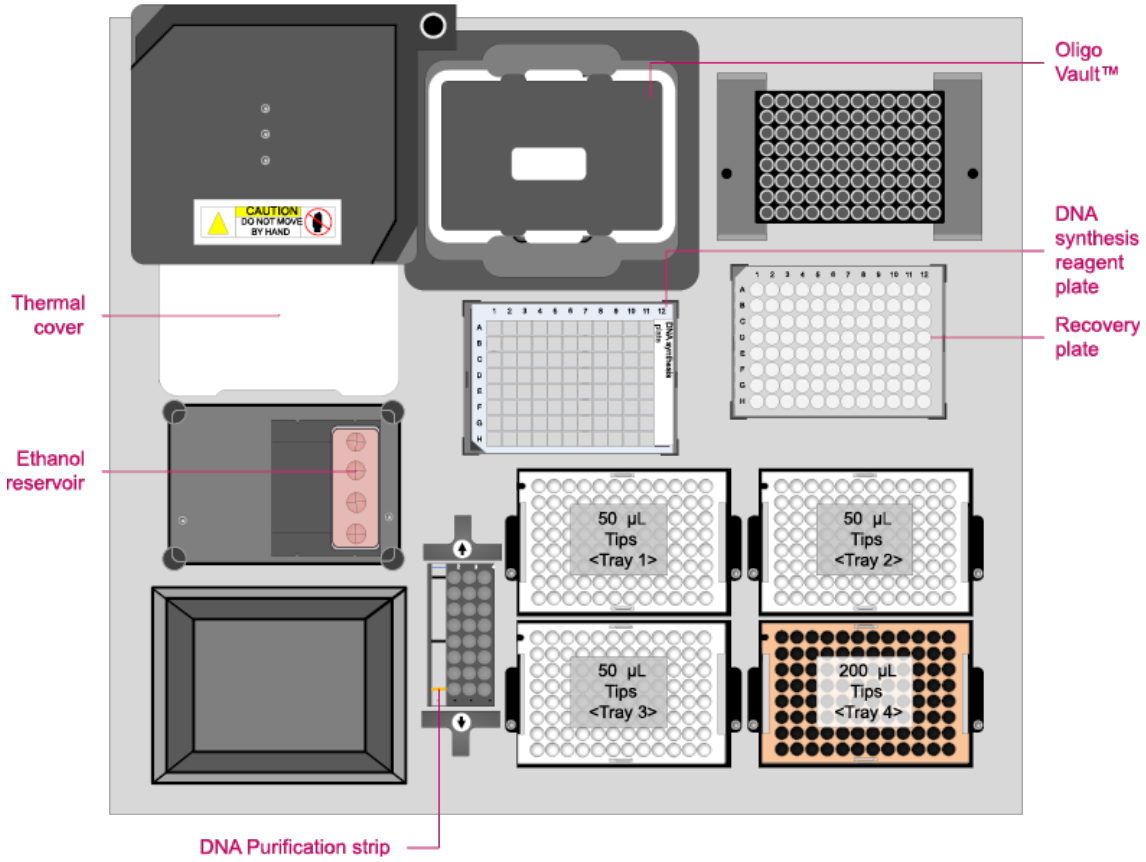


Figure 4: Image of the instrument deck with the components labeled.

Thawing and preparation of reagents

Thaw the DNA synthesis reagent plate (stored at -20°C) at room temperature (+25°C) for 30 minutes prior to placement on the BioXp system deck. Freshly prepare 12ml of 70% ethanol using nuclease-free water and add it to the designated reservoir before the run. Load ethanol reservoir in the rightmost reservoir retainer position of the instrument deck. Do not discard the ethanol reservoir after the run as this is a reusable unit and will be needed for future runs.

Loading the BioXp system with the BioXp ECL synthesis kit

Loading of reagents and consumables:

1. Load the recovery plate onto the recovery chiller with the notch in the upper left corner. Ensure a snug fit of the recovery plate. If it does not fit right, the orientation of the plate may not be correct.
2. Load the Oligo Vault™ plate so that the notch is positioned in the upper left corner of the thermocycler. Similar to the recovery plate, the Oligo Vault™ has directionality. If it does not fit right, the orientation of the plate is likely incorrect.
3. Follow these steps to load the appropriate purification strips depending on the reaction size of your kit (8, 16, 24 or 32):
    a. For an 8-reaction run, load the (orange) BX2300-08 BioXp DNA purification strip into position 1.
    b. For a 16-reaction run, load the (green) BX2300-16 BioXp DNA purification strip into position 1.
    c. For a 24-reaction run, load the (purple) BX2300-24 BioXp DNA purification strip into position 1.
    d. For a 32-reaction run, load the (gold) BX2300-32 BioXp DNA purification strip into position 1.

Note: Do not forget to secure the spring-loaded arms while holding the strips securely in place.

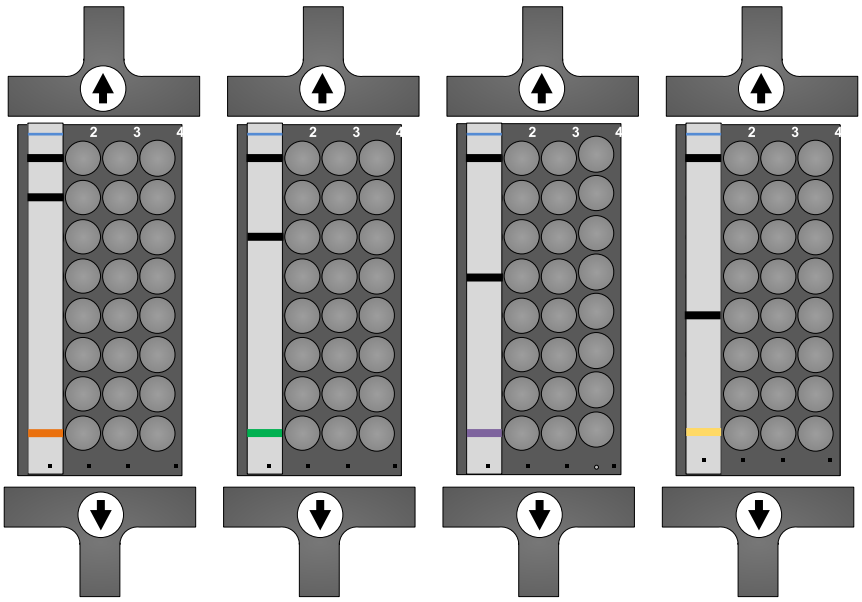


Figure 5: Purification strip types for ECL synthesis on the BioXp system. From left to right, the images show the correct ECL purification strip type for the 8, 16, 24 and 32 reaction runs.

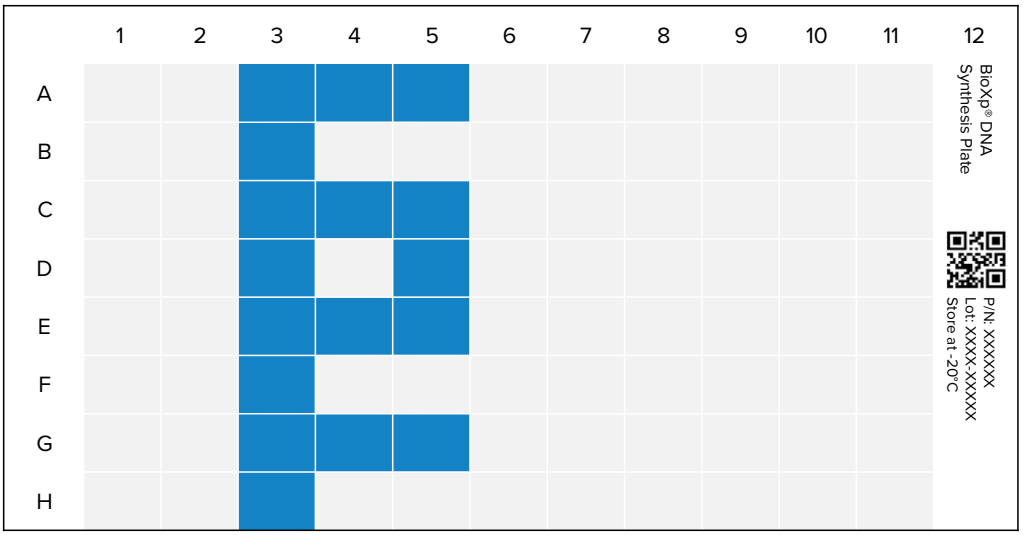


Figure 6: Visual representation of DNA Synthesis Reagent Plate. Wells with liquid are identified and marked in Blue.

4. Spin the thawed BioXp DNA synthesis reagent plate for one minute at 500 *rcf*. Visually inspect the wells to ensure they are completely thawed before loading the plate. Load the BioXp DNA synthesis reagent plate onto the reagent chiller, with notch in the lower left corner and barcode on the right. Figure 6 identifies the wells that contain reagent on the DNA synthesis reagent plate. Please ensure that the plate is properly seated within the chiller.
5. Add freshly prepared 70% ethanol into the ethanol reservoir and load the reservoir at the rightmost reservoir retainer position of the instrument deck.
6. Load tips by aligning the tip tray notch with the upper left corner of each tip tray retainer. Please see Table 1 for guidance regarding the tip tray needs for each reaction size. Please note that similar to the reagent plate, the tip trays have directionality. If the tip tray does not snap into place, it is not placed correctly. Rotate the tray 180° and try to place the tips again. The tip tray should have a snug fit and not wiggle.

Reaction Size	Number of 50µl tip trays	Number of 200µl tip trays	Tip Tray positions
8	1	1	1 and 4
16	2	1	1, 2 and 4
24	2	1	1, 2 and 4
32	3	1	1, 2, 3 and 4

Table 1: Tip Requirements based on the reaction size of the BioXp ECL synthesis run. Note that 200 µl tip tray needs to be loaded at Tip Tray position 4 regardless of the reaction size and the 50µl tip trays need to be loaded on to Tip Tray positions 1-3 depending on the reaction size.

Initiating the BioXp run for ECL synthesis

Prior to initiating the run, refer to the Loading Map document to ensure that all reagents and consumables have been correctly placed onto the BioXp system deck. Once confirmed, close the system door and press Start to begin the run. The typical run time is about 12.5hrs.

Retrieving your mRNA

Once the run is complete, the BioXp touchscreen will indicate that the run is done, see Figure 6 for reference.

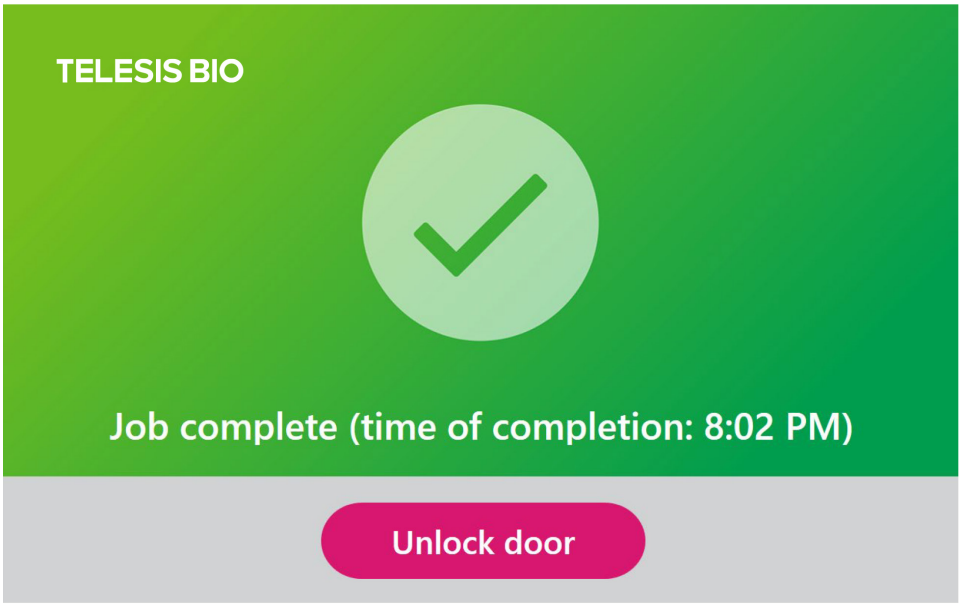


Figure 7: User screen on the BioXp 3250 when the ECL synthesis run is completed.

The newly synthesized ECLs are located within the first four columns of the product recovery plate.

- For an 8-reaction run, the ECLs are located in the first column of the product recovery plate (A1-H1).
- For a 16-reaction run, the ECLs are located in the first two columns of the product recovery plate (A1-H2).
- For a 24-reaction run, the ECLs are located in the first three columns of the product recovery plate (A1-H3).
- For a 32-reaction run, the ECLs are located in the first three columns of the product recovery plate (A1-H4).

Figure 8: ECL products are in the first four columns (A1 – H4) of the product recovery plate at the end of a 32- reaction run.

Post- BioXp synthesis run Quality Check of ECLs

Prior to using the ECLs for any downstream applications, it is highly recommended that the successful synthesis of the ECLs is evaluated by the following methods:

a. Resolving ECLs using an Agarose Gel. TapeStation and BioAnalyzer could also be used at this step. 2-3 µl of each sample is sufficient for this analysis on a 1-2% E-gel. This will provide information on the successful synthesis of the intended product in addition to the presence of any secondary products.

b. Measuring concentration using fluorescence (for example, Qubit) is recommended.

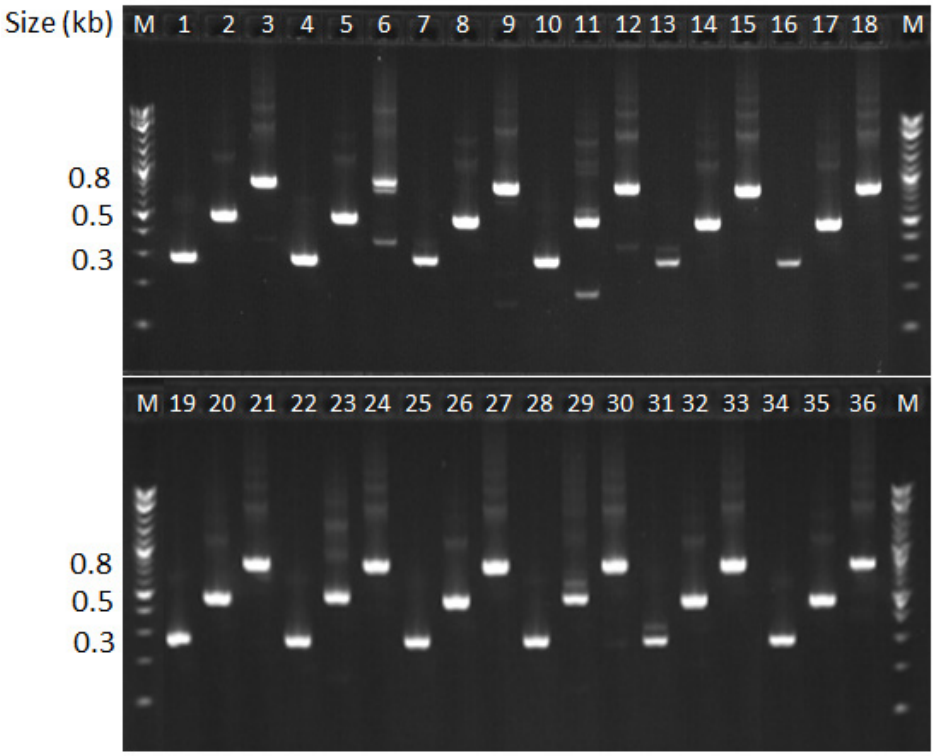


Figure 9: An example gel demonstrating ECL builds across various lengths, GC% and complexity. 2µl of the ECL built on the BioXp system was resolved using an E-gel (2% agarose). For ECL details, see Table 1.

Library no #	Size (bp)	GC %	Type	Sub-Type	Number of Degenerate bases	Degenerate base description
1	300	33.9	Scanning	Non-Contiguous	6	NNK,NNS
2	500	48.7	Scanning	Non-Contiguous	6	NNK,NNS
3	800	56.2	Scanning	Non-Contiguous	9	NNK,NNS,NNK
4	300	34	Scanning	Contiguous	3	NNN
5	500	48.5	Scanning	Contiguous	3	NNK
6	800	56	Scanning	Contiguous	3	NNK
7	300	33.7	Scanning	Contiguous	6	NNN,NNN
8	500	48	Scanning	Contiguous	6	NNK,NNK
9	800	58.2	Scanning	Contiguous	6	NNK,NNS
10	300	32	Scanning	Contiguous	9	NNN,NNN,NNN
11	500	48.9	Scanning	Contiguous	9	NNK,NNK,NNK
12	800	56.3	Scanning	Contiguous	9	NNN,NNK,NNS
13	300	34.2	Scanning	Non-Contiguous	9	NNN,NNK,NNS
14	500	46.7	Scanning	Non-Contiguous	9	NNN,NNK,NNS
15	800	57.4	Scanning	Non-Contiguous	6	NNS,NNK
16	300	34.1	Scanning	Non-Contiguous	12	NNK,NNS,NNS,NNK
17	500	45.1	Scanning	Non-Contiguous	12	NNK,NNK,NNK,NNK
18	800	56.3	Scanning	Non-Contiguous	12	NNK,NNS,NNN,NNK
19	300	34.2	Combinatorial	Contiguous	3	RKW
20	500	50.1	Combinatorial	Contiguous	3	HBD
21	800	55.8	Combinatorial	Contiguous	3	SVN
22	300	35.5	Combinatorial	Contiguous	4	NNKR
23	500	49.8	Combinatorial	Non-Contiguous	8	NNK,NNK,KH
24	800	55.8	Combinatorial	Non-Contiguous	10	NNN,MV,NNK,DV
25	300	30.1	Combinatorial	Non-Contiguous	2	M,W
26	500	47.2	Combinatorial	Non-Contiguous	2	B,D
27	800	56.2	Combinatorial	Non-Contiguous	2	S,V
28	300	33.9	Combinatorial	Non-Contiguous	3	K,M,S
29	500	48.7	Combinatorial	Non-Contiguous	3	B,D,H
30	800	60.3	Combinatorial	Non-Contiguous	3	W,H,N
31	300	35.4	Combinatorial	Non-Contiguous	4	K,D,W,V
32	500	47.6	Combinatorial	Non-Contiguous	4	D,H,H,B
33	800	56.2	Combinatorial	Non-Contiguous	4	N,S,V,H
34	300	32.5	Combinatorial	Non-Contiguous	6	B,Y,W,V,H,R
35	500	49.2	Combinatorial	Non-Contiguous	6	H,H,V,B,D,D
36	800	58.9	Combinatorial	Non-Contiguous	6	M,V,N,H,V,B

Table 2: ECL Details (Size, GC %, type, number of degenerate bases and type of degenerate bases incorporated).

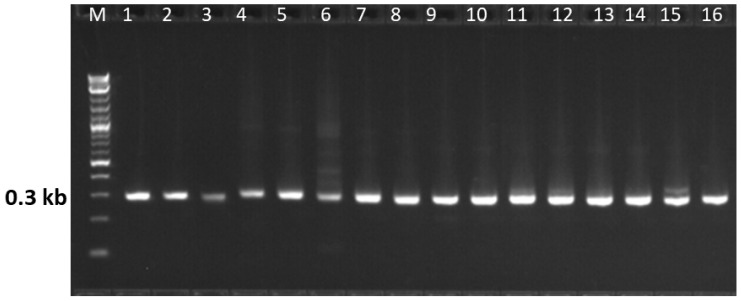


Figure 10: ECL builds across GC% ranging from 20-70% for a 300bp target sequence is shown. Details about the ECLs are listed in Table #2.

Library no #	Size (bp)	GC %	Type	Sub-Type	Number of Degenerate bases	Degenerate base description
1	300	30	Scanning	Contiguous	9	NNK,NNK,NNK
2	300	40	Scanning	Contiguous	9	NNK,NNK,NNK
3	300	50	Scanning	Contiguous	9	NNK,NNK,NNK
4	300	60	Scanning	Contiguous	9	NNK,NNKNNK
5	300	70	Scanning	Contiguous	9	NNK,NNK,NNK
6	300	20	Scanning	Contiguous	3	NNN
7	300	25	Scanning	Contiguous	3	NNN
8	300	30	Scanning	Contiguous	3	NNN
9	300	35	Scanning	Contiguous	3	NNN
10	300	40	Scanning	Contiguous	3	NNN
11	300	45	Scanning	Contiguous	3	NNN
12	300	50	Scanning	Contiguous	3	NNN
13	300	55	Scanning	Contiguous	3	NNN
14	300	60	Scanning	Contiguous	3	NNN
15	300	65	Scanning	Contiguous	3	NNN
16	300	70	Scanning	Contiguous	3	NNN

Table 3: 16 DNA library details (Size, GC %, type, no of degenerate bases).

Sequencing QC for ECLs

Fidelity and diversity can be assessed. We recommend assessing fidelity of the non-degenerate sequence of the ECLs using Sanger sequencing and diversity using NGS.

Sanger Sequencing – Fidelity Analysis

ECL products synthesized using the BioXp system are compatible with downstream cloning methods such as Gibson Assembly and TOPO. For Gibson Assembly recommendations, please refer to the **cloning guide**. For TOPO cloning, we recommend Zero Blunt™ TOPO™ PCR Cloning Kit (Thermofisher Scientific, Catalog – 450245) and transformation into chemically competent DH5α

E. coli cells (Thermofisher Scientific, Catalog – EC0112).

We recommend using 1 – 2µl (50 – 75 ng) of ECL (BioXp output) and follow manufacturer's protocol for TOPO cloning.

Note: Before sending samples for sequencing, we recommend performing colony PCR using M13 Forward & M13 Reverse (supplied with the TOPO cloning kit) to verify successful cloning. These colonies can be further used for Sanger sequencing analysis.

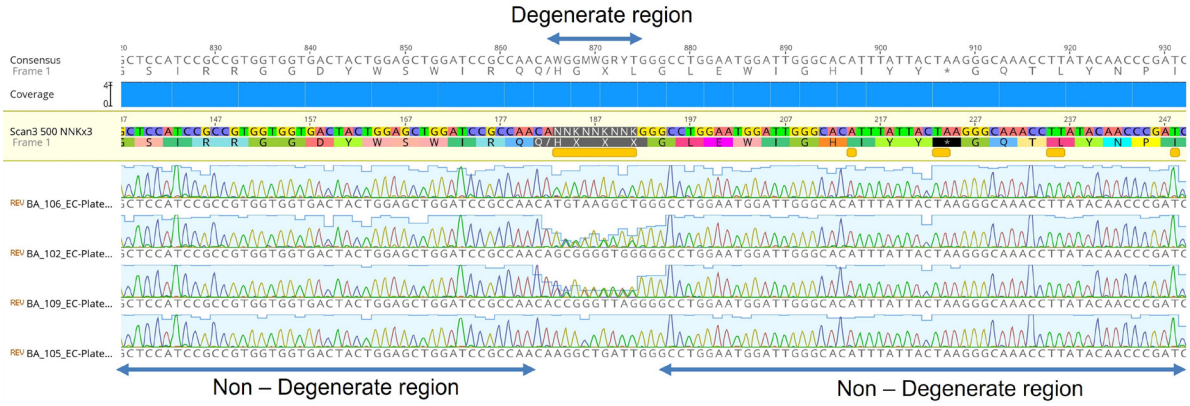


Figure 11: Typical Sanger Sequencing result. In this example, an ECL (500bp, contiguous NNK (3) codons). Degenerate regions show different nucleotides added in each clone, as expected. Non-degenerate region has no errors and matches the reference sequence.

NGS for Diversity Analysis

To confirm the expected codon/amino acid diversity at each degenerate codon position, we recommend performing NGS. Error corrected DNA Libraries for sequencing can be prepared & sequenced using:

- **Illumina DNA Library Prep**
- **MiSeq v2 Reagent Kit** (a higher yield kit could be used depending on the User Needs)

NGS - Recommendations

1. Library preparation should be performed based on manufacturer's protocol.
2. Use 50ng per DNA library sample (sizes 500bp –800bp) as input DNA for library preparation.
3. When pooling samples using same set of indices, it is recommended that the ECLs of similar size be pooled. However, it is also recommended that the ECLs pooled are diverse with respect to sequences. Please ensure that the pooled ECLs have less than 20% similarity at 5' & 3' ends from one another.
4. We recommend using the Qubit (High-sensitivity Kit; Thermofisher Scientific, Catalog # Q33230) for measuring the concentration of the pooled NGS libraries.
5. We recommend diluting the library to 2nM in 20µL of 10mM Tris-Cl, pH 8.5 with 0.1% Tween 20. Use the formula below to convert ng/µl to nM:

$$\frac{X \ \frac{ng}{ul}}{3.3 \ \frac{ng}{uL}} \times 10 = \sim Y \ nM \ DNA$$

Sequencing Recommendations

MiSeq Reagent Kit v2 supports 6 –10pM loading concentration in a final volume of 600ul.

- For the best results freshly prepare 0.2N NaOH. To compensate for low library diversity spike in PhiX control ≥ 5%.

NGS Data analysis recommendations

For the NGS analysis to determine diversity of ECLs at the degenerate positions synthesized on the BioXp system, it is necessary to filter out poor performing **reads** prior to aligning the **reads** to the reference sequence. In this case, it will be the file uploaded the myBioXperience project. After alignment, sort the reads and then use **bcftools** to annotate and call each position, while keeping the degenerate position in the reference file.

Filter out the positions with a quality score < 25, to measure the allele frequencies of the degenerate regions and compare against the expected frequencies. This analysis can be further extended to measure the expected vs observed codon distribution.

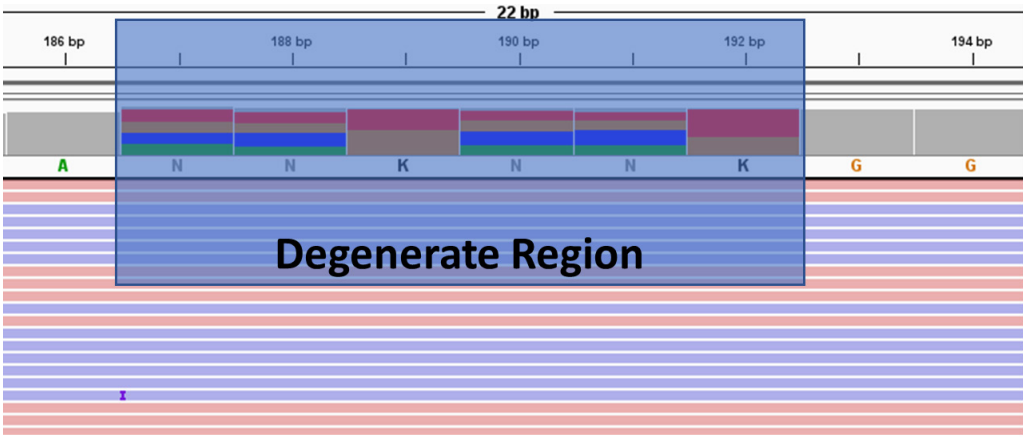


Figure 12: A graphical representation of reads aligned against an ECL reference sequence. Degenerate regions calls all four nucleotides in the first two positions (N) and two nucleotides in the third position (K).

To calculate the amino acid distribution, create a dictionary of the nucleotide-amino acid relation. Create a counting list that counts each amino acid for each degenerate region in the sequence file. Knowing the position of the degenerate region, filter out reads that do not contain the degenerate region of interest, and have a quality score less than 25. For each read, identify the codon that corresponds to the degenerate region, then increment the list in the counting list. After parsing each read, the amino acid frequency can be calculated as the count of a particular amino acid divided by the count of all amino acids.

Sanger sequencing recommendations

The same process for NGS can be repeated for Sanger Sequencing, however the allele frequency for degenerate positions is not recommended for Sanger Sequencing. If necessary to verify via Sanger Sequencing, sequencing redundancy must be implemented to account for the distributions of bases at degenerate positions.

Troubleshooting Recommendations

Problem	Cause	Proposed Solution
Additional bands are observed during ECL synthesis	DNA assembly includes off-target product	• Gel-purify full-length DNA if off-target product will affect downstream work • Redesign through codon-optimization to avoid off-target products. contact help@telesisbio.com for design assistance
ECL synthesis yields are low	DNA assembly was not efficient	Re-design construct; contact help@telesisbio.com for support
NGS assay issues	Low/No Cluster generation	• Ensure DNA sample is pure (Absorbance (A260/280) using spectrophotometer ~1.8 • Reduce input DNA amount during library preparation to 30 - 50ng per ECL sample • Dilute pooled library to 1.5nM to 2nM before denaturing with NaOH • Load 6pM – 10pM for sequencing
NGS assay issues	Low diversity in ECL sample	• Spike in PhiX control ≥ 5% of the input material • Avoid pooling ECL samples with >20% similarity at 5' & 3' ends within the same indices

Table 4 : Describes recommendations for troubleshooting when you have a problem with the intended use of the BioXp ECL synthesis kit.

Telesis Bio
9535 Waples Street, Suite 100
San Diego, CA 92121-2993 USA

858.228.4115

Technical assistance:
help@telesisbio.com

TelesisBio