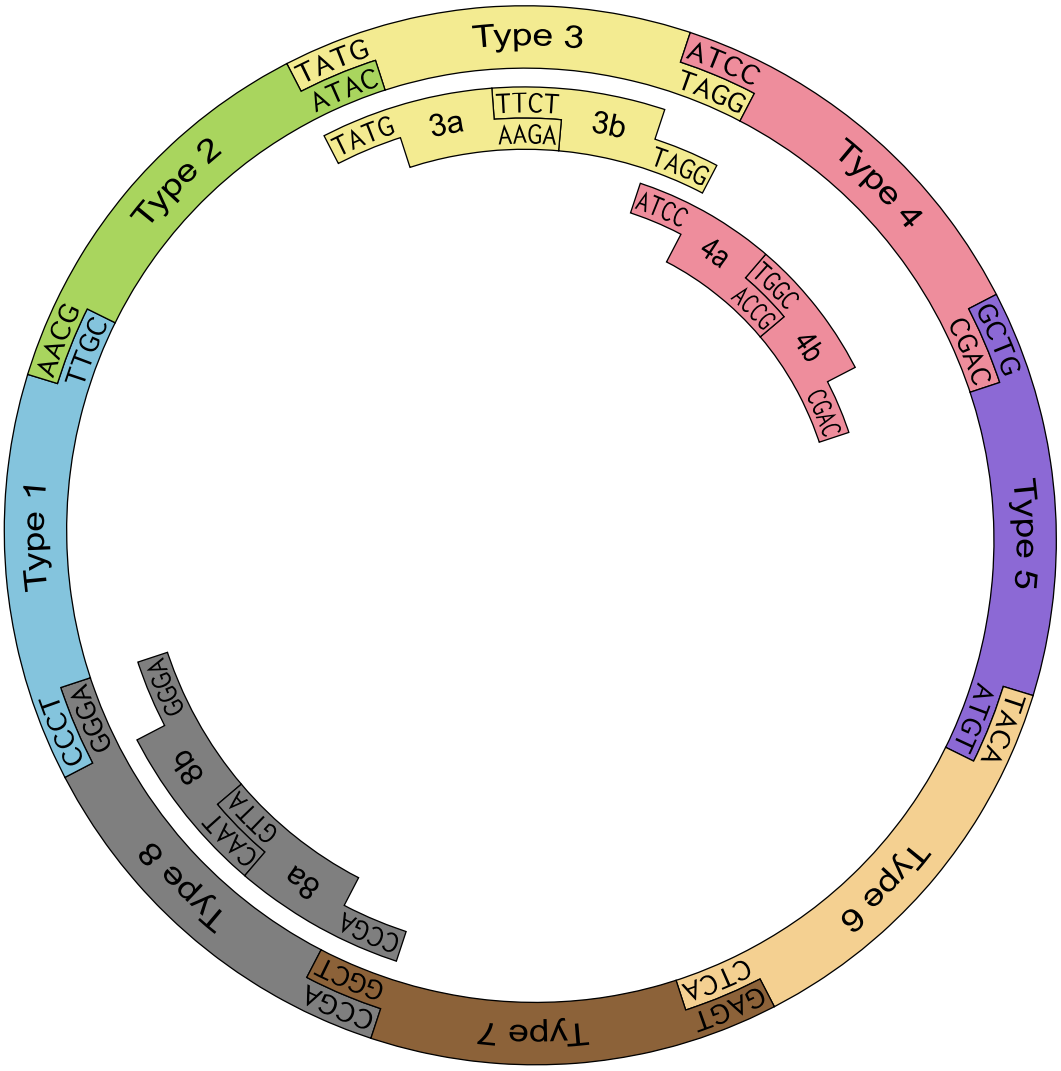
SUPPLEMENTARY TEXT

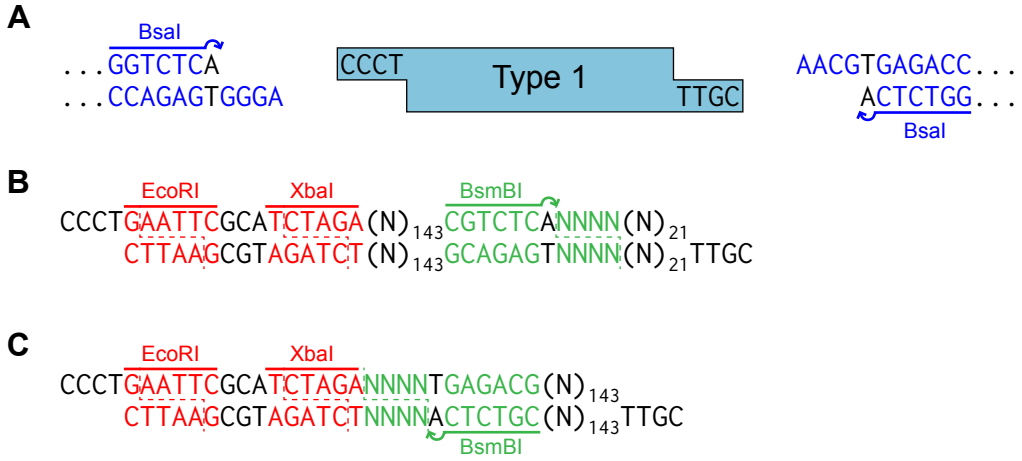
Definition of Part Types

There are eight primary part Types in our assembly standard and three of those have options to split into Subtypes. It should be noted that Types are technically defined only by their flanking overhangs, and the contents need not necessarily match the biologically-defined functions we describe (**Supplementary Figure S6**). However, in order to ensure compatibility between all parts designed by all researchers, we recommend that new parts be designed to match the Types defined here.



Supplementary Figure S6

Type 1: 5′ Assembly Connector



Supplementary Figure S7

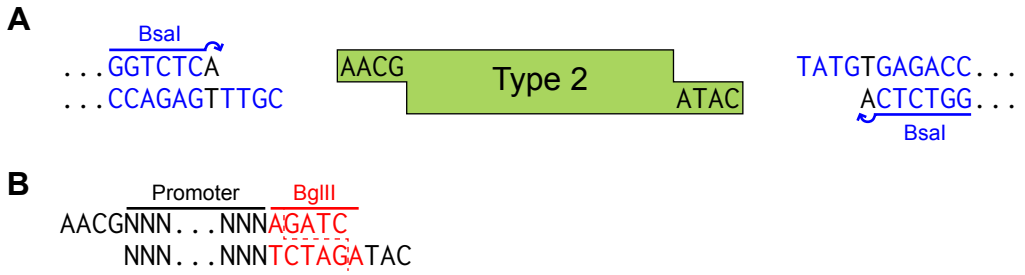
Type 1 parts are flanked by CCCT and AACG (**Supplementary Figure S7A**). Typically, this Type contains non-coding, non-regulatory sequences that are used to direct assembly of multi-gene plasmids.

The Type 1 part plasmids included in the toolkit contain a 143bp concatenation of barcode sequences used in the systematic deletion of yeast genes[1], a BsmBI recognition site and unique overhang, and a 21bp barcode scar (again, from the systematic deletion collection) (**Supplementary Figure S7B**). We designate these sequences as Assembly Connectors, and the nomenclature used is "ConLX" where X = 1, 2, 3, etc. The BsmBI site is oriented such that the restriction enzyme digests the sequence downstream of the recognition sequence.

There are also ConLX′ parts, where the structure instead contains first a reversed BsmBI site (digests upstream) followed by the 143bp sequence (and no barcode scar) (**Supplementary Figure S7C**). The purpose of this alternate Assembly Connector is for generation of multi-gene backbone plasmids (see the detailed description of assembly scheme below). To simplify the toolkit, we include a single reversed Assembly Connector, which we designate ConLS′ and its cognate forward version, ConLS (S for Start), although any numbered Assembly Connector can have a reversed version if desired.

Finally, the Type 1 parts in this toolkit also include an EcoRI and XbaI site for BioBrick compatibility of the assembled cassettes and multi-gene plasmids.

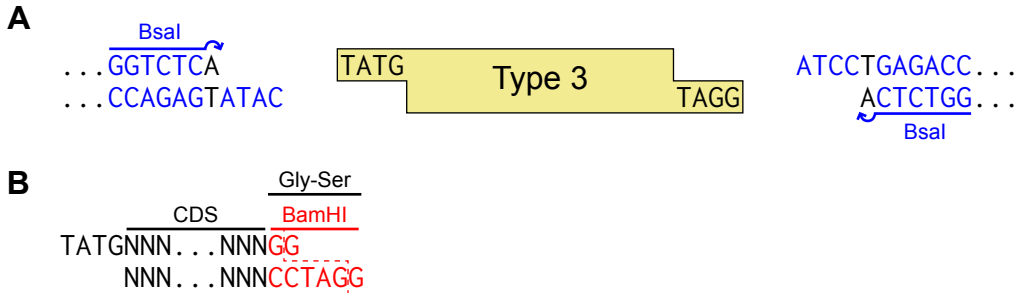
Type 2: Promoter



Supplementary Figure S8

Type 2 parts are flanked by AACG and TATG (**Supplementary Figure S8A**). Typically, this Type contains a promoter. The downstream overhang doubles as the start codon for the subsequent Type 3 or 3a coding sequence. Additionally, all the promoters in this toolkit have a BglII site immediately preceding the start codon (overlapping the downstream overhang) for BglBrick compatibility (**Supplementary Figure S8B**).

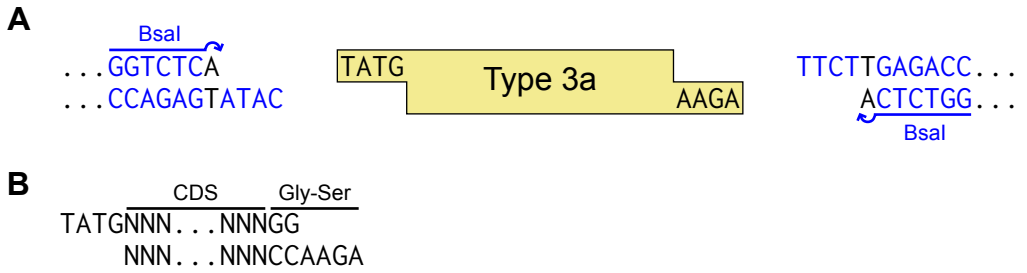
Type 3: Coding Sequence



Supplementary Figure S9

Type 3 parts are flanked by TATG and ATCC (**Supplementary Figure S9A**). Typically, this Type contains a coding sequence. As discussed above, the TATG overhang includes a start codon so coding sequences should begin with the second codon. The ATCC overhang was designed to enable read-through for protein fusions. If a stop codon is omitted from the part, and two bases are added before the overhang, the resulting NNATCC can be used as a two amino acid linker to a Type 4 or 4a C-terminal fusion. The Type 3 parts in this toolkit all omit the stop codon and add a GG, resulting in GGATCC, which serves a dual purpose. First, the resulting Gly-Ser linker is relatively innocuous; and second, the sequence is a BamHI recognition site, which enables BglBrick compatibility (**Supplementary Figure S9B**). We highly recommend following this convention unless the protein in question is sensitive to C-terminal modifications.

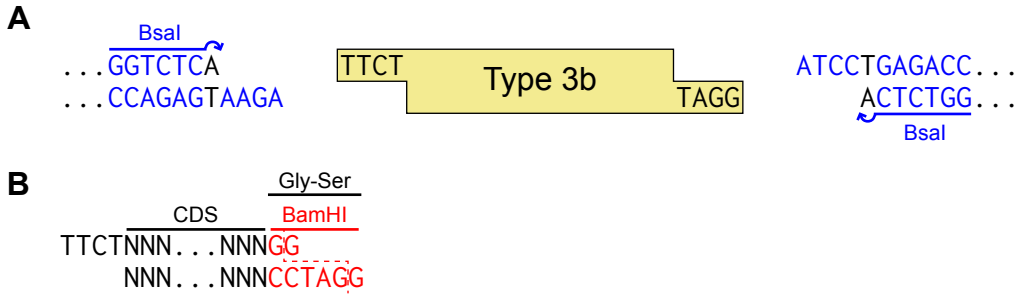
Type 3a: N-terminal Coding Sequence



Supplementary Figure S10

Type 3 parts can be split into 3a and 3b parts for greater flexibility for making protein fusions. Type 3a parts are flanked by TATG and TTCT (**Supplementary Figure S10A**). As with Type 3 parts, these typically contain coding sequences, and can be used for fusing N-terminal tags (such as the degradation tags described in the main text). Again, as with Type 3 parts, the stop codon should be omitted and two bases should be added before the TTCT overhang if protein fusions are desired. The Type 3a parts in this toolkit add a GG, resulting in GGTTCT, another Gly-Ser linker (**Supplementary Figure S10B**).

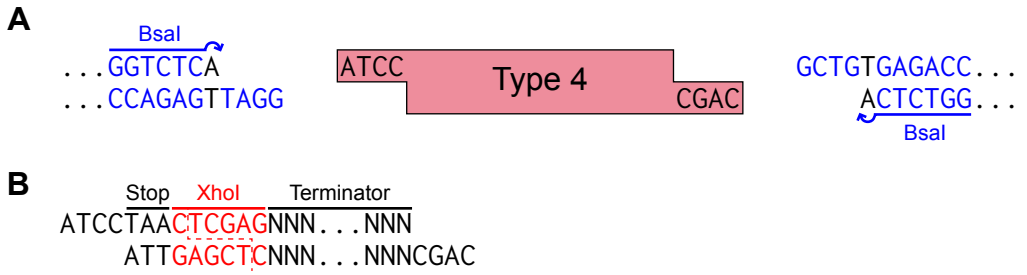
Type 3b: Coding Sequence



Supplementary Figure S11

Type 3b parts are flanked by TTCT and ATCC (**Supplementary Figure S11A**). As with Type 3 and 3a parts, these typically contain coding sequences. Again, the start codon should be removed for direct fusions to the Type 3a preceding it, and two bases should be added before the ATCC overhang if C-terminal fusions are desired. As with the Type 3 parts, all Type 3b parts in this toolkit add a GG, resulting in GGATCC (Gly-Ser/BamHI site) (**Supplementary Figure S11B**).

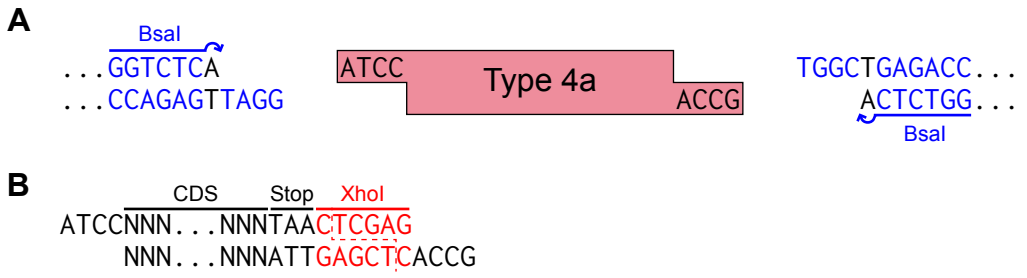
Type 4: Terminator



Supplementary Figure S12

Type 4 parts are flanked by ATCC and GCTG (**Supplementary Figure S12A**). Typically, this Type contains a transcriptional terminator. As described above, the convention for a Type 3 or 3b is to omit the stop codon and allow read-through of a GGATCC (Gly-Ser) linker. Therefore, the Type 4 should encode an in-frame stop codon before the transcriptional terminator. The Type 4 parts in this toolkit begin with a TAA stop codon, followed by a XhoI site (CTCGAG, for BglBrick compatibility), then the terminator sequence (**Supplementary Figure S12B**). Commonly used C-terminal fusions, such as purification or epitope tags, may be in included before the stop codon, but we recommend using the 4a/4b subtypes to maintain their modularity.

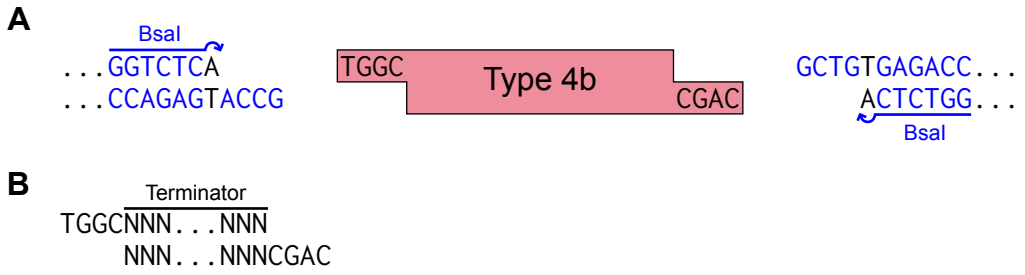
Type 4a: C-terminal Coding Sequence



Supplementary Figure S13

Like Type 3 parts, Type 4 parts can be split into 4a and 4b parts for additional modularity. Type 4a parts are flanked by ATCC and TGGC (**Supplementary Figure S13A**). Typically, this Type contains a coding sequence for fusing to the C-terminus of a protein (such as a localization tag, fluorescent protein, or purification tag). However, in contrast to the Type 3 and 3b parts, the convention for 4a parts is to include the stop codon rather than enable read-through of the TGGC overhang (although this is possible if desired). As such, the Type 4a parts in this toolkit end with a TAA stop codon, followed by a XhoI site (CTCGAG, for BglBrick compatibility) (**Supplementary Figure S13B**).

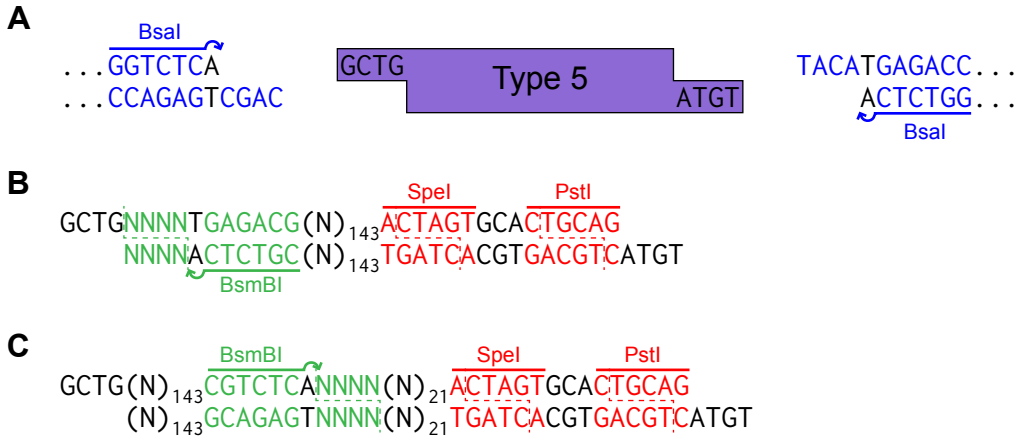
Type 4b: Terminator



Supplementary Figure S14

Type 4b parts are flanked by TGGC and GCTG (**Supplementary Figure S14A**). As with Type 4 parts, these typically contain transcriptional terminators (**Supplementary Figure S14B**). Because the convention of a Type 4a part is to encode the stop codon, one is not necessary in a 4b and so only the terminator sequence is needed.

Type 5: 3′ Assembly Connector



Supplementary Figure S15

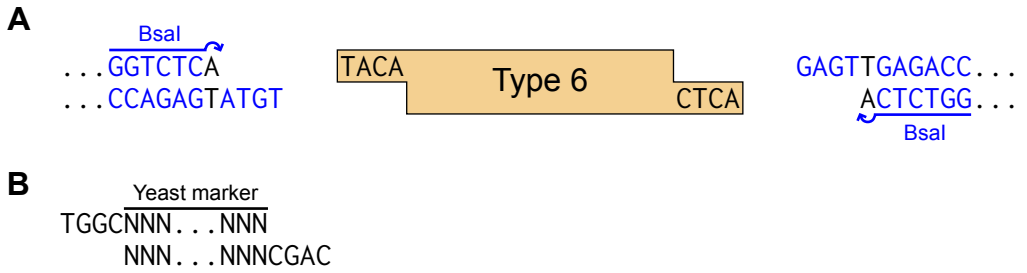
Type 5 parts are flanked by GCTG and TACA (**Supplementary Figure S15A**). As with the Type 1 parts, these parts typically contain Assembly Connectors.

The structure of Type 5 parts is very similar to that of Type 1 parts. First is an upstream-cutting BsmBI site (with a unique overhang), followed by a 143bp concatenated barcode sequence (this structure is identical to that of the Type 1 ConLX′) (**Supplementary Figure S15B**). Here, the nomenclature used is "ConRX". Again, there is a special structure for ConRX′ parts: the 143bp sequence, a downstream-cutting BsmBI site, and a 20bp barcode (this structure is identical to that of the Type 1 ConLX) (**Supplementary Figure S15C**). We included in this toolkit, a single ConRE′ (E for end) part and its cognate forward version, ConRE.

The key to the Type 1 and 5 Assembly Connectors is that the unique overhangs generated by BsmBI digestion should match for parts with the same value of X. For example, the BsmBI overhang generated by ConL1 and by ConR1 is CCAA. This is critical for enabling assembly of multi-gene plasmids, which is described in detail below.

Finally, the Type 5 parts in this toolkit also include a SpeI (ACTAGT) and PstI (CTGCAG) site for BioBrick compatibility of the assembled cassettes and multi-gene plasmids.

Type 6: Yeast Marker

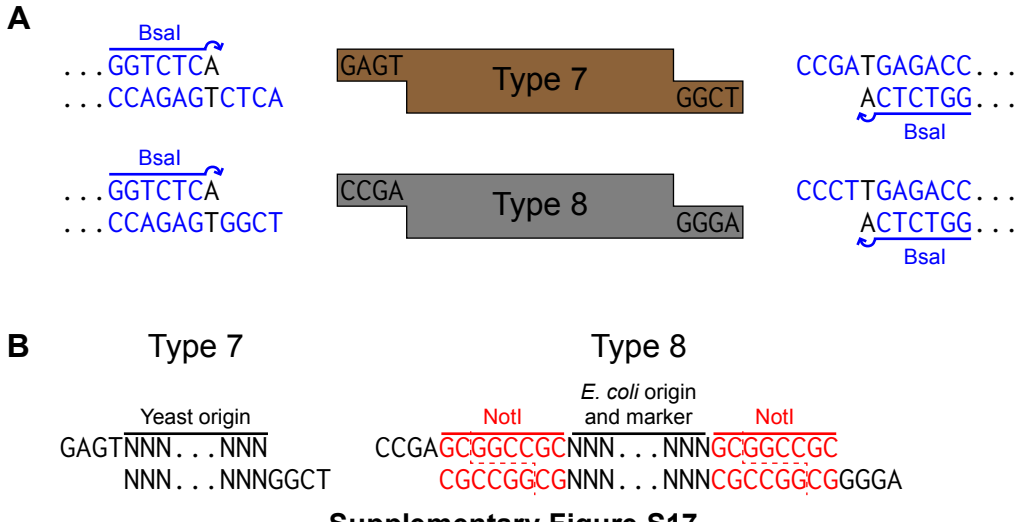


Supplementary Figure S16

Type 6 parts are flanked by TACA and GAGT (**Supplementary Figure S16A**). Typically, this Type contains a selectable marker for *S. cerevisiae*. These parts should include the full expression cassette (promoter, ORF, and terminator) for conferring the selectable phenotype (usually amino acid prototrophy or drug-resistance) (**Supplementary Figure S16B**).

Type 7+8: Yeast Plasmid Propagation

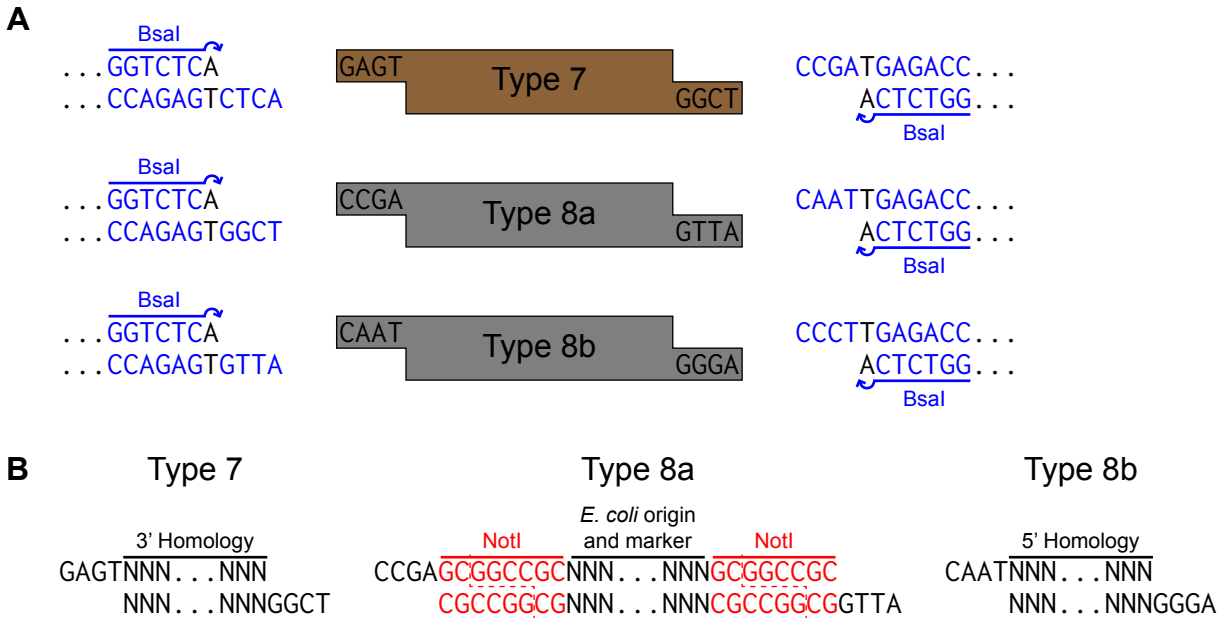
Type 7 and 8 parts can be used in two ways, depending on the application. For plasmid expression in yeast, Type 7 and 8 parts should be used as described in this section; for integration into the yeast chromosome, Type 7, 8a, and 8b parts should be used as described in the next section.



Supplementary Figure S17

Type 7 parts are flanked by GAGT and CCGA (**Supplementary Figure S17A**). For propagation of a stable plasmid in yeast, this Type contains a yeast origin of replication (**Supplementary Figure S17B**). Type 8 parts are flanked by CCGA and CCCT as well as NotI sites that are useful for restriction mapping to verify new assemblies (**Supplementary Figure S17A**). This Type contains a bacterial origin of replication and antibiotic resistance marker (**Supplementary Figure S17B**).

Type 7+8a+8b: Yeast Chromosomal Integration



Supplementary Figure S18

For integration into the yeast chromosome, the Type 7 parts (which retain the GAGT and CCGA overhangs) contain sequences that have homology that is downstream (3′) of the target locus (**Supplementary Figure S18A**). Longer homology sequences are more efficient at recombining into the chromosome; therefore, the parts in this toolkit contain 500bp of homology. Additionally, a 20bp barcode sequence is included upstream of the homology region to serve as a forward primer binding site for colony PCR verification of integration into the correct locus.

Type 8a parts are flanked by CCGA and CAAT (**Supplementary Figure S18A**). As with the Type 8 parts, these typically contain a bacterial origin of replication and antibiotic resistance marker (**Supplementary Figure S17B**). These parts are also flanked by NotI sites that can be used to linearize the integration plasmid prior to transformation into yeast (as well as for restriction mapping).

Type 8b parts are flanked by CAAT and CCCT (**Supplementary Figure S18A**). Similar to Type 7 homology parts, these parts contain long sequences of homology to the genome that is upstream (5′) of the target locus (**Supplementary Figure S18B**). Additionally, a 20bp barcode sequence is included downstream of the homology region to serve as a reverse primer binding site for colony PCR verification of integration into the correct locus.

Miscellaneous

In addition to these standard part Types, non-standard Types that span two or more positions can be constructed and are conventionally named as a concatenation of the Type numbers spanned. For example, some cassette plasmids are constructed only as intermediates toward a multi-gene plasmid. These cassettes no longer require any of the yeast maintenance machinery (origin and marker) and so a Type 678 part that only contains a bacterial origin and marker may be appropriate to use.

Detailed Description of Hierarchical Assembly System

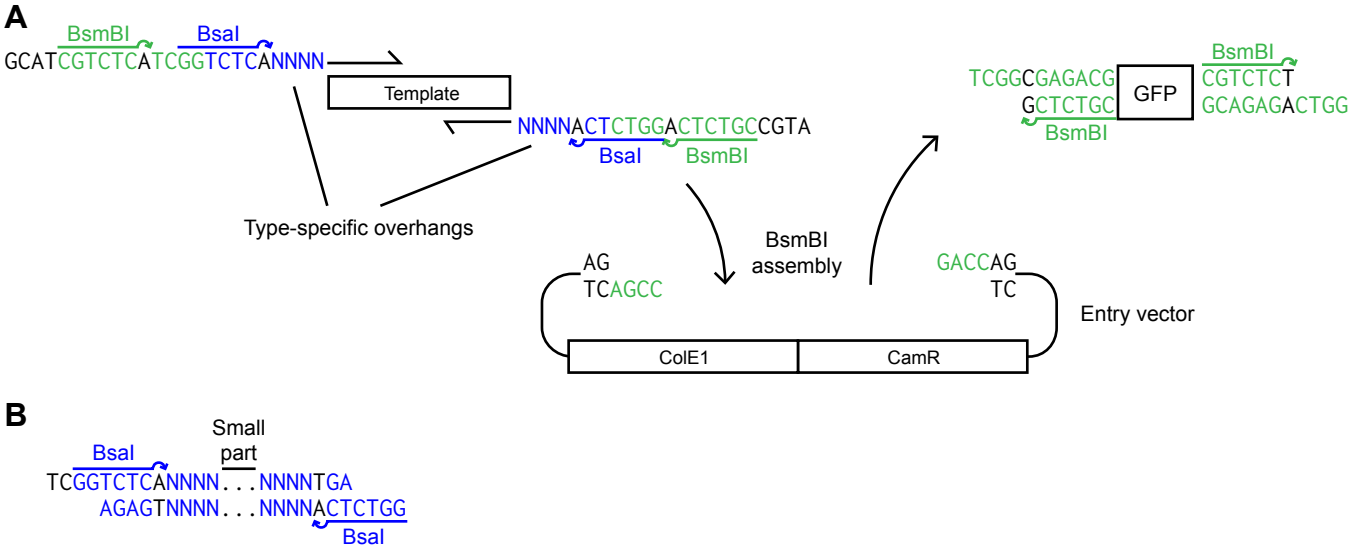
The construction of plasmids encoding multiple transcriptional units is done in three steps. The first is construction of part plasmids; second is assembly of cassette plasmids; and third is assembly of multi-gene plasmids (**Figure 1**).

Construction of Part Plasmids

The general structure of a part plasmid is as follows: 1) a downstream-facing BsaI site that generates the upstream flanking overhang of the part Type; 2) the part sequence; 3) an upstream-facing BsaI site that generates the downstream flanking overhang of the part Type; and 4) a ColE1 origin of replication and chloramphenicol resistance marker. Detailed descriptions for each part Type can be found in the Definition of Part Types.

Part plasmids are assembled via a BsmBI Golden Gate reaction into the part entry vector (**Supplementary Figure S19A**). The entry vector contains a ColE1 origin of replication and chloramphenicol resistance marker, as well as a GFP expression dropout for green/white screening. BsaI, BsmBI, and NotI sites should be removed from all parts except in special cases. Additional restriction sites such as BbsI or the BioBrick/BglBrick enzymes may also be removed, but it is not necessary unless future use of those enzymes is anticipated.

Primers for amplifying preexisting templates should be designed as illustrated in **Supplementary Figure S19A** to enable BsmBI assembly into the entry vector, and subsequent BsaI cassette assemblies. The four N's flanking the part should correspond to the flanking overhangs for the specific part Type (e.g., AACG and TATG for a Type 2). Modifications to the part sequence (e.g. restriction site removal) can be easily introduced by dividing the part into multiple DNA inserts in the BsmBI Golden Gate reaction. Internal overhangs in this reaction can be user-selected, but should avoid similarity to the entry vector overhangs TCGG and GACC. Parts made from de novo synthesis should mimic the same structure or be ordered in the entry vector. Finally, small parts can be assembled from overlapping oligonucleotides that drop directly into the entry vector (**Supplementary Figure S19B**). We routinely add annealed oligonucleotides and/or synthesized gene fragments (e.g. gBlocks®) directly to the BsmBI Golden Gate reaction.

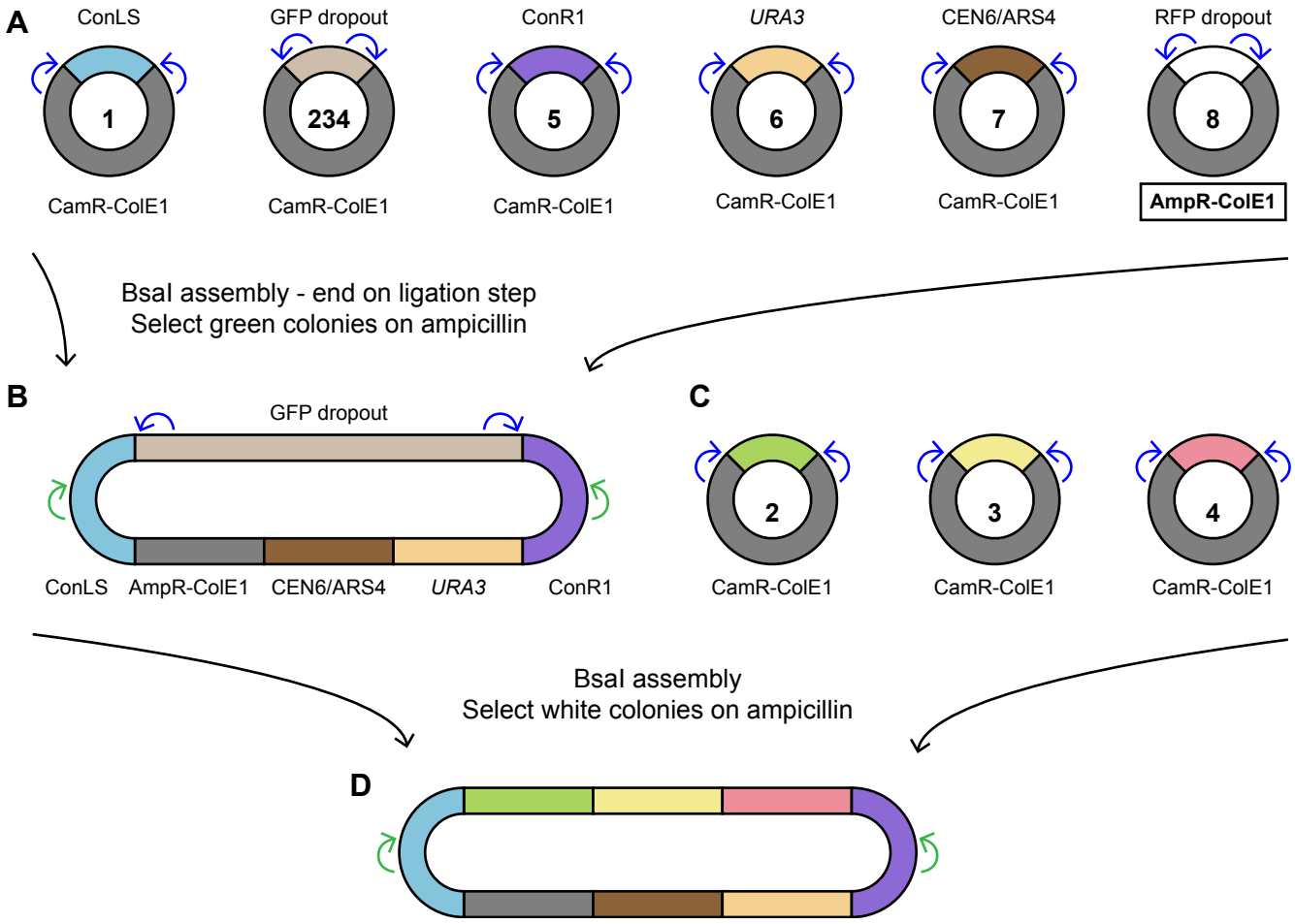


Supplementary Figure S19. Construction of new part plasmids.

A special exception must be made for constructing new Assembly Connectors (Type 1 and 5). In these cases, the Assembly Connectors contain internal BsmBI sites used in multi-gene assemblies. There are two options for constructing these part plasmids. First, primers can be designed as normal, but the Golden Gate assembly protocol should be modified to exclude the final digestion and heat inactivation steps, thereby ending on a ligation. Second, an existing Type 1 or Type 5 plasmid can be digested and gel purified using BsaI, and the new part can then be assembled in using BsaI rather than BsmBI.

Assembly of Cassette Plasmids

The simplest way to assemble a cassette is to include one part of each Type in a BsaI assembly. The Type 8 and 8a parts included in this toolkit serve as the canonical "vectors", and accordingly have an mRFP1 expression dropout for red/white screening.



Supplementary Figure S20. Assembly of cassette plasmids.

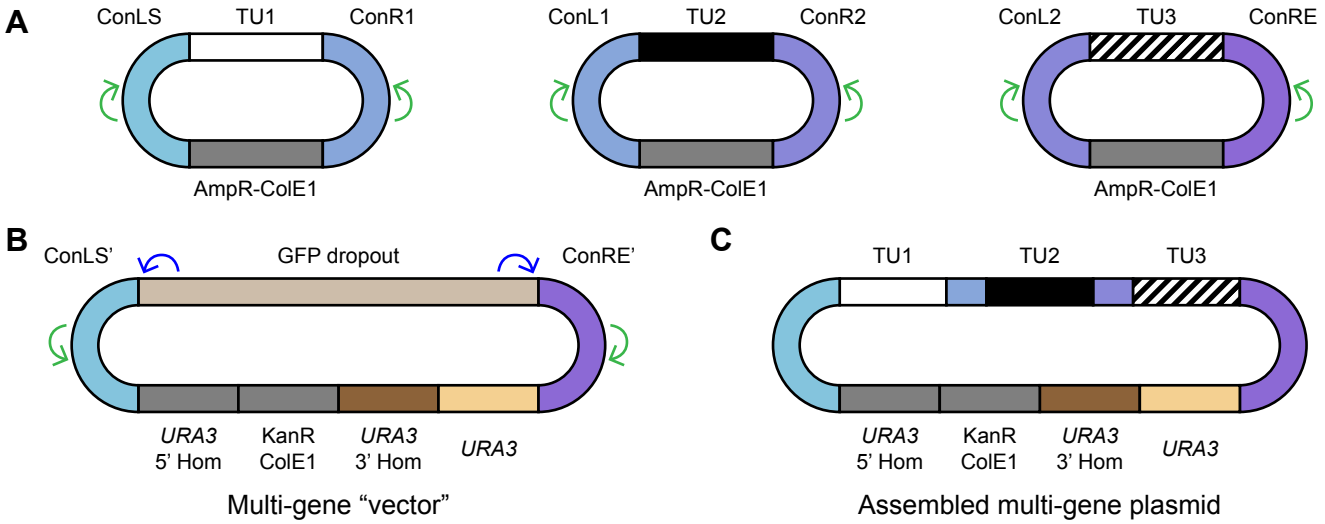
An alternative approach is to pre-assemble commonly used parts with a GFP dropout that spans the variable region. For example, a Type 234 GFP dropout part is included in this toolkit. By assembling ConLS, the GFP dropout, ConR1, URA3, CEN6/ARS4, and AmpR-ColE1, a cassette "vector" can be made (**Supplementary Figure S20A and B**). By storing this cassette, any future assemblies of transcriptional units (promoter, coding sequence, terminator) with these vector components will require fewer parts, something that is particularly useful for the generation of combinatorial libraries (**Supplementary Figure S20C and D**). It is important to note

that the Type 234 GFP dropout part has BsaI sites in the reverse orientation that normal parts do and will remain in the finished vector. Therefore, the Golden Gate assembly protocol should be modified to exclude the final digestion and heat inactivation steps, thereby ending on a ligation. We have observed a significantly higher rate of misassembly for this procedure (~50%). Incorrect products are typically concatenations of part plasmids and contain multiple origins of replication and antibiotics markers. Wrong products can be easily identified because they will confer growth in media with either chloramphenicol or the desired antibiotic, whereas the correct product will not confer growth in media with chloramphenicol.

Assembly of Multi-Gene Plasmids

The construction of a multi-gene plasmid from cassettes requires that the cassettes are flanked by unique pairs of Assembly Connectors, which dictate the order of assembly. The first cassette must contain the ConLS part, and the last cassette must contain the ConRE part. The order of internal Assembly Connectors can be arbitrary, although going in increasing numerical order is recommended to avoid confusion. Thus, before the individual cassettes are made, the structure of the final multi-gene plasmid should be designed because it will determine which Assembly Connectors should be used during cassette assembly.

For example, if three transcriptional units, TU1, TU2, and TU3 are to be assembled into a multi-gene plasmid in that order, one possible design would be to flank TU1 with ConLS and ConR1, TU2 with ConL1 and ConR2, and TU3 with ConL2 and ConRE (**Supplementary Figure S21A**). The "vector" into which these cassettes are assembled is itself another cassette, which uses the special ConLS' and ConRE' parts (**Supplementary Figure S21B**). One example of this special cassette is included in this toolkit, designed to target the *URA3* locus for integration. Again, in assembling this "vector" cassette, the Type 234 GFP dropout can be used to enable green/white screening. When the final multi-gene plasmid is assembled, the Assembly Connector junctions leave behind 20bp barcode "scars", which can be used to verify proper assembly by colony PCR or sequencing (**Supplementary Figure S21C**).



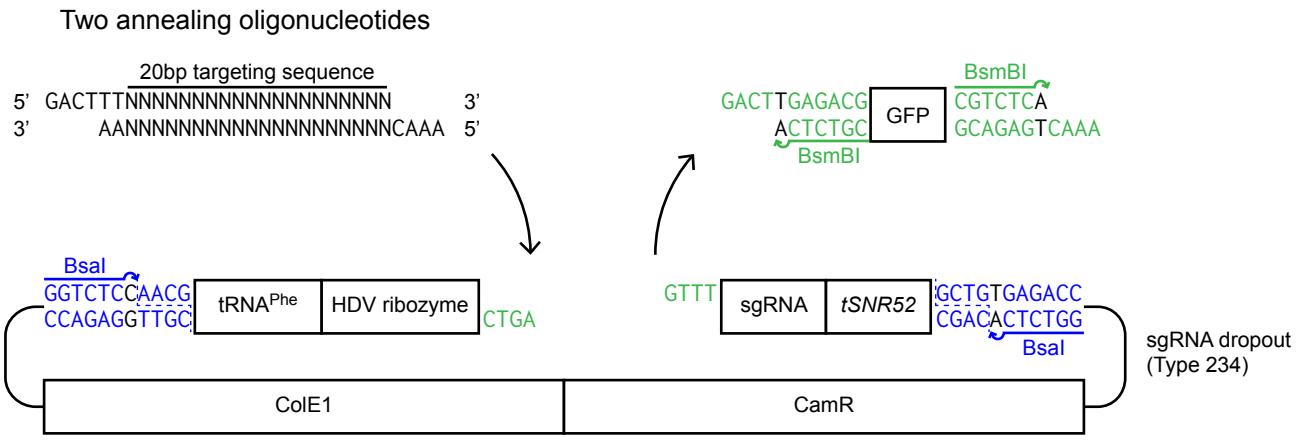
Supplementary Figure S21. Assembly of multi-gene plasmids.

One issue with this style of assembly is that cassettes are locked into their position based on the Assembly Connectors flanking them. For example, in the same three-TU multi-gene plasmid described above, if TU2 were to be omitted, there would be a gap that would require reas-

sembling either TU1 or TU3 to replace the right or left Assembly Connectors, respectively. One solution to this we provide in the toolkit is the Type 234 "Spacer" part. This part can be used to assemble filler cassettes—in this case, a ConL1-Spacer-ConR2 cassette. The advantage of assembling a filler cassette rather than a reassembled TU cassette is that the filler can be used again in future assemblies when that gap needs to be filled.

Construction of CRISPR/Cas9 sgRNAs

To construct sgRNAs for targeting Cas9 to a site in the genome, we have provided a Type 234 sgRNA dropout, which is effectively an entry vector for these parts. This vector is based on the CRISPRm sgRNA architecture: a phenylalanine tRNA, a HDV ribozyme, a 20bp targeting sequence, the sgRNA, and an *SNR52* terminator. For the dropout, the targeting sequence is replaced by a BsmBI-flanked GFP expression cassette. Unfortunately, the 4bp immediately upstream and downstream of the targeting sequence are CTTT and GTTT, which could incorrectly ligate, so the upstream overhang was moved two bases upstream to GACT. Consequently, two additional T's should be added before the targeting sequence when ordering oligonucleotides to anneal and ligate into the dropout (**Supplementary Figure S22**). Once constructed, the sgRNA 234 part can be assembled into a cassette with appropriate connectors. This cassette should then be assembled into a multi-gene plasmid that also includes Cas9 expression and additional sgRNAs (optional).



Supplementary Figure S22. Construction of CRISPR/Cas9 sgRNAs.

Differences from MoClo

The hierarchical assembly system described in this work borrows heavily from the MoClo[2] system with some modifications.

First, the specific overhang sequences that flank the parts and the cassettes are different from those used by the MoClo creators. This change was made to support a greater number of part Types as well as the in-frame protein fusions enabled by several of those Types. Additionally, in an attempt to minimize assembly errors due to misligation of incompatible overhangs, we tried to find a more optimal set. It has been previously reported that when three out of four contiguous nucleotides match between two overhangs, a misligation event can occur (e.g., ATCG and ATCA)[3]. Furthermore, we had observed that even three non-contiguous matches could result in a misligation event (e.g., ATCG and ATAG, or ATCG and ATGA). Therefore, we tried to find a set of overhangs where the fewest such matches were present.

Second, the MoClo system requires an extensive series of vectors to support the various possibilities of multi-gene assemblies. Rather than creating the exhaustive set of possible vectors up front, our system utilizes the Assembly Connector parts to enable on-the-fly construction of vectors. One advantage of this approach is that rather than defining transcriptional units as being in Position 1, 2, 3…, they are defined as being between Assembly Connectors X and Y. Thus, a transcriptional unit could be cloned between ConL1 and ConR4, or ConL3 and ConR2, as long as the final sequence begins with ConLS and ends with ConRE and has no repeated Assembly Connectors.

Finally, MoClo utilizes three Type IIs restriction enzymes to enable indefinite assembly of multiple transcriptional units. As described, our system is limited to one round of multi-gene assembly, although it could be extended to include this added functionality if desired. We have removed a third Type IIs restriction site (BbsI) from all parts in this toolkit for such purposes.

Alternative Assembly Methods

Although Golden Gate is the preferred assembly method to be used in this system, there are a number of alternative methods that can be used at some steps of the process.

The initial part plasmid construction can be performed using any method, as long as the resulting plasmid has the appropriate BsaI overhangs flanking it.

The cassette plasmid assembly must be performed using Golden Gate. Other methods such as Gibson or SLIC can be used, but they will require unique primers for every new junction. Only Golden Gate assembly will preserve modularity at this step.

Once a cassette plasmid has been assembled, there is much more flexibility in terms of downstream assembly steps. If all the conventions described above are followed, the cassettes will be flanked by BioBrick restriction enzyme sites, enabling BioBrick cloning of cassettes with each other, or with existing BioBrick plasmids that have not been converted to this new system. Second, cassettes also contain BglBrick restriction enzyme sites that flank the coding sequence, enabling BglBrick cloning for fusing coding sequences with existing BglBrick parts. Third, the purpose of the 143bp sequences in the Assembly Connector parts is to facilitate modular recombination-based assembly methods, such as Gibson, SLIC, or *in vivo* yeast assembly. As with the BsmBI overhangs in the Assembly Connectors, the 143bp sequences of ConLX and ConRX parts with the same value of X will be exactly the same, so cassettes can be designed in the same way for both Golden Gate and recombination-based assembly. Note that the final sequence of the multi-gene plasmid will be different depending on which method (Golden Gate, BioBrick, or recombination) is used for the assembly.

References

(1) Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M., and Davis, R. W. (1996) Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular barcoding strategy. *Nat Genet 14*, 450–456.
(2) Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* (Peccoud, J., Ed.) *6*, e16765.
(3) Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS ONE* (Peccoud, J., Ed.) *4*, e5553.