Overview of Golden Gate cloning

(NEB.com and others have nice video tutorials about Golden Gate cloning.) Golden Gate cloning uses Type IIS restriction enzymes which cleave DNA to one side from their recognition sequences. The overhangs generated by BsaI (GGTCTC(1/5)) or BsmBI (CGTCTC(1/5)) can be any four bases allowing them to be designed to be compatible with some and not with other sites.

Example, two BsaI sites in opposite orientations that create the same overhang:

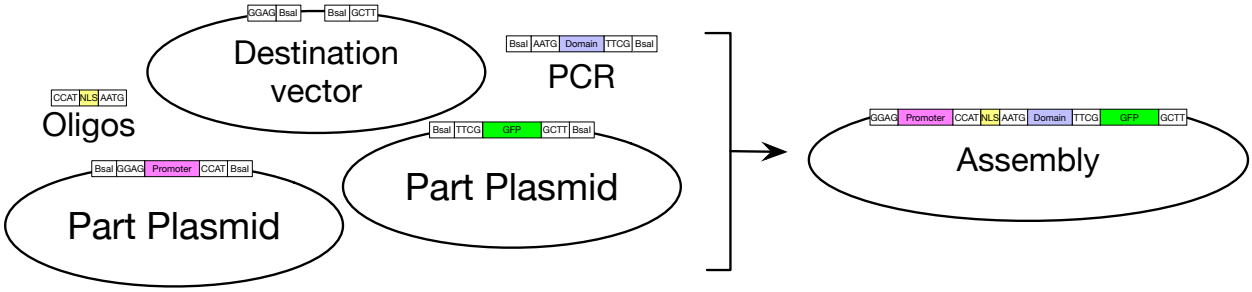
```
                    ↓      BsaI                      BsaI ↓
5’…agcttgcatgcgtcgacggagcgagaccggag…3'  5’…tagtggtctcgggagataatgagcattgcatg…3’
3’…tcgaacgtacgcagctgcctcgctctggcctc…5’  3'…atcaccagagccctctattactcgtaacgtac…5’
                         ↑                          ↑
```

Products after repeated digestion/ligation cycles:

```
…agcttgcatgcgtcgacggagataatgagcattgcatg…
…tcgaacgtacgcagctgcctctattactcgtaacgtac…

   ggagcgagaccggag…  …tagtggtctcg
       gctctggcctc…  …atcaccagagcctc
```

Whenever the overhang connected to a BsaI recognition sequence is ligated to something, it is re-cleaved the next cycle while those without BsaI sites are not. By mixing fragments with strategically designed BsaI sites plus the BsaI enzyme and DNA Ligase and then cycling between temperatures favoring digestion and ligation, it is possible to assemble the fragments in the desired order to create circular plasmids (lacking BsaI sites). Typically, we assemble 3-5 fragments into a destination vector, but it is possible to correctly assemble dozens of fragments.



The big advantages of Golden Gate cloning are that the reactions are very efficient, very cheap (<$1/reaction), and very flexible in that it allows mixing & matching construct components. The main drawback is that fragments to be assembled should not contain internal restriction sites for the enzyme being used.

My system was designed to generate plasmids for the types of transformations our lab tends to do most often (tagging endogenous gene and inserting overexpression/ reporter transgenes into neutral-targeting loci) in one assembly step. Iterative systems such as GoldenBraid, MoClo, and Loop assembly systems might offer more flexibility (like also allowing mixing and matching selectable markers and targeting arms). My parts plasmids (Level 0 plasmids) are mostly based on the standard syntax and would be compatible with those systems. The vectors confer chloramphenicol (moss) or kanamycin (Arabidopsis) resistance, while the parts plasmids confer amp resistance.

How I assemble constructs using Golden Gate

1. Choose a destination vector based on the type of transformation:

A. Insertion into one of the *PIG1*, *Pp108*, or *ARPC2a* neutral-targeting loci
Use pMP2130/2133/2134 (*PIG1* locus, G418/Kan selection), pMP2123 (*ARPC2a* locus, Gent/G418 selection), or pMP2132/2135 (*Pp108* locus, Zeo selection). Different terminator sequence options are available for *PIG1* and *Pp108*. (The selection cassettes in *Pp108* and *ARPC2a* could be swapped for Hyg, Gent/G418, or Zeo [as BstXI - SpeI/SacI fragments] if necessary.)

B. Scarless gene tagging (i.e., homology arm plasmid for CRISPR-facilitated insertion/deletion without integration of a selection cassette)*
pMP2136 (G418/Kan), pMP2137 (Hyg), pMP2138 (Gent/G418), pMP2139 (Zeo).

C. Plasmids for transient gene expression. Same vectors as for B ↑

D. Homology arm plasmid for knock-in/replacement with an integrated selection cassette* Use pMP2121 or NEB's pGGAselect. Note that I have not created any Golden-Gate-ready tag/terminator/selection plasmids for this type of transformation, but I have a plan to do so if needed.

E. Binary plasmid for Arabidopsis transformation.
Use pMP2200. (The *Basta* selection cassette could be swapped with a different one as a HindIII fragment if Basta is problematic.)

* Note on choosing between transient and stable selection when tagging genes (with GFP, epitopes, etc.): I have multiple examples of the *35S* promoter in the selection cassettes causing ectopic expression of tagged genes (that reverted to normal upon CRE-mediated selection cassette removal). I have usually decided that I'd rather genotype more progeny after a single transformation than transform twice (first to integrate, then to remove the selection cassette). In addition, N-terminal tags are much easier if only the tag is integrated. All that said, it is easier to obtain the initial tagged lines when selecting for lines stably resistant to antibiotics. ~50% stably resistant lines have clean integrations compared to 5-10% for scarless-tag insertions.

2. Decide what components you will want in the final construct and diagram it. (See the diagrams in "MossGG" PDF for common examples.)
For example, if you want to overexpress a gene tagged with mYPet (a bright monomeric YFP) from the *PIG1* neutral locus:
pMP2133 [GGAG] proZmUbi [AATG] Gene [TTCG] C-mYPet [GCTT]

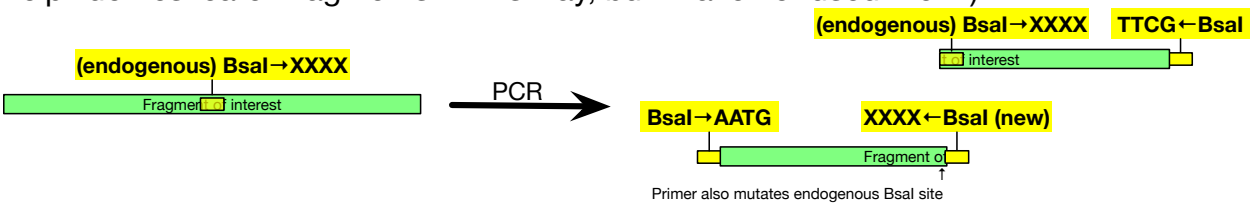
For tagging an endogenous gene with an N-terminal mYPet (transient G418 selection):
pMP2136 [GGAG] 5' UTR sequence [CCAT] N-mYPet [AATG] Seq after ATG [GCTT]

The first and last overhangs are dictated by the destination vector, and most of the rest follow those of existing parts plasmids which follow a standard syntax (see diagram in MossGG PDF). The remaining overhangs can be anything not already used (but avoid palindromic overhangs). Pryor et al 2020 (NEB) determined that the 9 best overhangs to go with the 11 standard ones are ACCT, CCGC, ACAA, AACA, GAAA, CAAG, GCAC, TAGA, and AAAT.

3. Check that the fragments do not contain internal BsaI (or BsmBI) sites.

The destination vectors are designed for use with either BsaI or BsmBI assembly reactions depending on which enzyme site is absent from fragments to be assembled. BsaI (NEB: BsaI-HFv2) is cheaper and more efficient than BsmBI (NEB: BsmBI-v2) so we have only used it so far. As a result, all of our existing parts plasmids were created for BsaI assemblies and would need to be modified for BsmBI assemblies.

There are a few options if a fragment has a BsaI (or BsmBI) site. You could mutate them by re-synthesizing it or by site-directed mutagenesis. If there is only one internal site in one of your fragments and the overhangs don't match any used in synthesis, you can go ahead and cycle it as normal and then add a tiny bit extra T4 DNA Ligase and incubate at 16°C for ~30 more minutes. (I was skeptical, but this works fairly well.) My normal approach is to amplify the fragment in two pieces (or three if there are two BsaI sites, etc.). For each internal BsaI site, identify the overhang it generates, and make a primer adding a BsaI recognition sequence on the opposite side from the endogenous one that would generate the same overhang and introduce a silent mutation to abolish the endogenous BsaI site. Make another primer somewhere behind the endogenous BsaI site. Depending on the BsaI site's orientation, one of these primers will be the reverse primer for amplifying the first fragment and the other will be the forward primer for amplifying the second fragment. (I believe there are websites to help "domesticate" fragments in this way, but I have not used them.)



Example (*bla*/AmpR):

```
                                  BsaI
WT  gataaatctggagccggtgagcgtggtctcgcggtatcattgcagcactggggccag
    D  K  S  G  A  G  E  R  G  S  R  G  I  I  A  A  L  G  P
Mut gataaatctggagccggtgagcgtgggtcAcgcggtatcattgcagcactggggccag
    D  K  S  G  A  G  E  R  G  S  R  G  I  I  A  A  L  G  P
    GGR primer  ← cggtgagcgtgggtcAcgcggtaGAGACCcagc
     ataaatctggagccggtgagc → Forward primer
```

bla1-GGR Primer (rev-comp): gctgGGTCTCtaccgcgTgacccacgctcaccg
bla2-F primer: ataaatctggagccggtgagc

4. Gather or make the components needed.

The components to be assembled can be supplied to Golden Gate reactions as 1) PCR products with BsaI (or BsmBI) sites that result in the proper 4-bp overhangs, 2) plasmids with these PCR products (including the BsaI/BsmBI sites) subcloned and sequenced, or 3) as annealed double-stranded oligos. Components derived from sequenced plasmids generally do not need to be sequenced again in the final construct, while those from PCR or oligos should be. My rule of thumb is to subclone and sequence any PCR product that will be used more than once (because the cost of a ¼-scale NEB PCR cloning kit reaction is about the same as a Sanger sequencing run).

We are accumulating a collection of sequenced plasmids with a few different promoters and three configurations each for several bright monomeric fluorescent-protein gene tags. (See “Parts Plasmids” section of MossGG PDF.)

Designing PCR primers:
Each primer will start with a few (2-6) random bases followed by a BsaI or BsmBI restriction site, GGTCTC or CGTCTC, another random base, the required overhang sequence (in forward primers) or the reverse-complement of the required overhang sequence (in reverse primers), then the sequence-specific portion for annealing during PCR.

When designing the sequence-specific portion, the reading frame is very important in certain components. Among the ‘standard’ overhangs, the AATG, CCAT(G), AGGT, AGCC, TTCG dictate the reading frames with the underlined bases forming codons that set the reading frame for the subsequent gene sequences.

Forward primers, BsaI	tagtGGTCTCg[XXXX][sequence-specific]
Reverse primers, BsaI	tagtGGTCTCg[rev-compYYYY][sequence-specific]

XXXX and YYYY are the corresponding 5′ and 3′ overhangs. The lowercase letters can be anything (except those that create *dcm* methylation sites, CC[A/T]GG). I use “tagt” at the start. For BsmBI primers, substitute CGTCTC for GGTCTC.

E.g., primers for the ZmUbi promoter flanked by BsaI sites (GGAG → AATG overhangs)

ZmUbi-GGF	tagtGGTCTCgGGAGATAATGAGCATTGCATGTCTAAG
ZmUbi-GGR	tagtGGTCTCgCATTACCTGCAGAAGTAACACCAAA

Subcloned components (‘Parts plasmids’ or ‘Level 0 plasmids”)
An ideal vector for subcloning parts would lack both BsaI and BsmBI sites and wouldn’t use kanamycin or chloramphenicol selection, but I am unaware of such a commercial PCR cloning kit. We use NEB’s PCR Cloning Kit because it is efficient and the pMiniT2 vector lacks BsaI sites which we use most (but has a BsmBI site with CTAT/ATAG overhangs). Promega’s pGEM-T/pGEM-T Easy vector has a BsaI site (CGGT/ACCG), and Thermo’s pJET1.2 has both BsaI (CGGT/ACCG) and BsmBI (TTTG/CAAA) sites. If these overhangs differ from those in your assemblies, they should not drastically affect efficiency.

Oligos
Double-stranded oligos with the desired overhangs can be synthesized and used directly—adding BsaI sites is not necessary. Protospacers for making sgRNAs are added as oligos into the CRISPR plasmids (see below), but I have also used oligos for adding a nuclear localization sequences, epitope tags, or as linkers to link components with non-compatible overhangs. Note that oligos lack 5′ phosphates unless ordered with that modification or treated with T4 polynucleotide kinase, but this is only necessary if adjacent components are both oligos.

Forward oligo	[XXXX][sequence-specific]
Reverse oligo	[rev-compYYYY][sequence-specific]

E.g., Oligos encoding the cMyc NLS (AATG → AGCC overhangs)

cMycNLS-ggF, AATGCCTGCTGCAAAGAGGGTGAAGCTGGATgg
cMycNLS-ggR, GGCTccATCCAGCTTCACCCTCTTTGCAGCAGG

```
   M  P  A  A  K  R  V  K  L  D  G  A
5'AATGCCTGCTGCAAAGAGGGTGAAGCTGGATggAGCC 3'
3'TTACGGACGACGTTTCTCCCACTTCGACCTAccTCGG 5'
```

5. (Optional) *in silico* Golden Gate Assembly test.
The free sequence editor, ApE, (and some commercial sequence editors) has a Golden Gate Assembler tool that can double-check your planned assembly before ordering primers and starting the benchwork. Open files for the destination vector and each component (you may need to add BsaI sites to the oligo components), and follow the dialog to complete the assembly. If there are problems with overhangs (or extra BsaI sites), it will tell you. Check the assembled sequence for proper reading frame, insertions/deletions, etc.

6. Golden Gate assembly reactions
I set up 10 µl reactions with 1× T4 DNA Ligase Buffer, 100 ng destination vector, two-fold molar equivalents of each insert part, 0.4 µl BsaI-HFv2, and 0.4 µl T4 DNA ligase. For oligos, I anneal them, dilute them, and add 0.25 pmol to the reactions (0.5 µl of 0.5 µM annealed oligos). This has been plenty efficient for us (up to 6 fragments), but NEB sells optimized enzyme mixes for BsaI-HFv2 (E1601) and BsmBI-v2 (E1602) that may improve efficiency of more complicated assemblies. I don't believe the precise concentrations are critical.

Cycling: I do 60 cycles of 1 min 37°C and 1 min 16°C followed by 5 min at 60°C. (For BsmBI-v2, use 42° instead of both 37°C and 60°C.) Fewer cycles (e.g., 30) is likely sufficient for most assemblies. Many protocols have longer incubations at the two temperatures (commonly 2 min 37°, 5 min 16° each cycle), but NEB's protocol uses 1 min except for when assembling 11 or more inserts, for which they suggest 5 minutes for both steps.

Transform 1-3 µl into comp cells and plate on LB-chloramphenicol plates (25 mg/l) or LB-kanamycin (50 mg/l) depending on the destination vector used. My destination vectors contain the *lacZ′-superfolderGFP* fusion, so it is possible to identify non-recombinant plasmids using either X-gal for blue/white screening or using a 470 nm LED flashlight and an amber filter to see fluorescent colonies. In practice, this is unnecessary as fewer than 1 in 1000 colonies are non-recombinant (unless something went horribly wrong).

I screen colonies by colony PCR using primers that amplify across assembled insert. If all the components were previously sequenced plasmids, I don't usually sequence the assembly. Segments derived from PCR products and oligos should be sequenced.

Golden Gate CRISPR vectors

Loading CRISPR plasmids to express 1 or 2 sgRNAs

The Bezanilla lab created a CRISPR/Cas9 vector system that uses paired BsaI sites to load protospacer oligos into pENTR plasmids between a PpU6 promoter and an sgRNA scaffold, and then these entry clones are recombined into a Gateway destination vectors containing ZmUbi:Cas9 & selectable markers (Mallett et al 2019). To skip the Gateway step, I recombined their empty pENTR-U6:2×BsaI-sgRNA into their pMH-Cas9-Gate then removed extra BsaI site in the AmpR gene by moving the entire insert to a plasmid backbone with Kan resistance. The resulting pMP2098 and pMP2141 vectors allow loading protospacers in a single step.

A. Choose the protospacer for your sgRNA. We use the http://crispor.tefor.net/ website. Paste in the sequence you want to target, select the "Physcomitrella … Phytozome V11" genome and "20bp-NGG - Sp Cas9…" PAM. The output will be ordered by specificity. You want your sgRNAs to be specific, to cut very close to the site you want edited, and to be efficient. The diagram with the candidate PAMs and dashes indicates the cut sites (after the last dash opposite the PAM). I don't believe either efficiency score is particularly accurate for plants but I weight Doench '16 over the other one. I ignore the "Outcome" columns unless I want to create short, deleterious indels. Protospacers that start with a "G" might express better with the U6 promoter. I've had decent sgRNAs that didn't start with a G, but my most efficient sgRNAs did.

Example (*PIG1* locus) 5′ homology arm ↓ 3′ homology arm
Sequence …tgagtgaaaagaattagaaattctcatgtatggttcactttat…
GAATTAGAAATTCTCATGTA*ngg*
———*Protospacer*———*PAM*

B. **Design protospacer oligos for single-sgRNA constructs.** The list shows the protospacer sequences followed by a space and the italicized PAM (e.g., GAATTAGAAATTCTCATGTA *TGG*). Copy the protospacer portion and precede it with "ccat" for the forward oligo. Paste the reverse-complement preceded with "aaac" for the reverse oligo.

Single sgRNA-For ccat*GAATTAGAAATTCTCATGTA* (*Ns* = protospacer)
Single sgRNA-Rev aaac*TACATGAGAATTTCTAATTC*

C. **Using bridge plasmids to express a second sgRNA.** The original Bezanilla lab vector used multisite Gateway reactions to combine up to 4 sgRNAs into a single CRISPR plasmid. For deleting genes, I like to use sgRNAs targeting both ends of the gene, so I made two 'bridge' plasmids to allow loading a second protospacer into CRISPR plasmids. They contain two BsaI sites flanking sequence for the sgRNA scaffold and terminator for the first sgRNA then a U6 promoter for the second sgRNA. To differentiate the BsaI overhangs while maintaining the same sequences flanking the protospacers, I shifted the overhangs over 1 or 2 bases relative to the single-sgRNA version. One version, pMP2170, requires that the first base of the second protospacer is the optimal "G" for a "CATG" overhang. Being a palindrome, two bridge plasmids can ligate together non-productively and lower

the yield of the desired construct. A second version, pMP2171, shifts that BsaI's overhang the other direction ("CCCA"). I have compared the two bridge plasmids side-by-side only once. As expected, I got about half as many colonies in the reaction using pMP2170, but still plenty. I PCR'd across the insertion site and cut with BsaI for 4 colonies each. For pMP2171, only 1 appeared to have both protospacers and the other 3 had only one (I'd guess protospacer 2, but I didn't check). For pMP2170, 2 appeared to have both protospacers, 1 was empty vector, and 1 did not amplify.

Dual sgRNA1-For	ccat*NNNNNNNNNNNNNNNNNNN*gt	*N*s = 20-base protospacer (**for both**)
Dual sgRNA1-Rev	taaaac*NNNNNNNNNNNNNNNNNNN*	*N*s = rev-comp of 20-base protospacer
Dual sgRNA2-For	cccat*NNNNNNNNNNNNNNNNNNN*	*N*s = 20-base protospacer (**for pMP2171**)
Dual sgRNA2-Rev	aaac*NNNNNNNNNNNNNNNNNNN*a	*N*s = rev-comp of 20-base protospacer
or		
Dual sgRNA2-For	cat*GNNNNNNNNNNNNNNNNNN*	*GN*s = 20-base protospacer (**for pMP2170**)
Dual sgRNA2-Rev	aaac*NNNNNNNNNNNNNNNNNN*	*N*s = rev-comp of protospacer bases 2-19

D. **Golden Gate reactions & colony screening.** I do 10 µl reactions, set up as outlined above for other Golden Gate reactions. (If a protospacer contains an internal BsaI site, cycle as normal then add more ligase and incubate at 16°C for a while.) After transformation, I colony PCR across the sgRNA locus using a reverse proZmUbi primer (aaaaaggagaacacatgcacact) and a reverse terNOS primer (ttgcgggactctaatcataaaaac) followed by a BsaI digest to distinguish desired clones from empty vector. [In the future, I might ligate an *eforCP* pink chromoprotein gene between the BsaI sites in the CRISPR vectors so that empty-vector colonies would be pink and the PCR product would be a different size.] I purify the rest of the PCR product of positives and sequence it with a PpU6-F primer (ggggagaaaagggatggagc). I have not had good luck sequencing with that primer off the plasmid directly. I don't think our lab has ever found a mutation in a protospacer (n>50), so sequencing may not be necessary. I have not attempted to sequence a dual-sgRNA construct given that the two sgRNA genes are identical except for the protospacer; I'm not sure how to do so other than whole-plasmid sequencing (eg, Plasmidsaurus or Primordium). I am considering swapping out the *Physcomitrium* U6 promoter in a bridge plasmid for the orthologous promoter from related moss, *Funaria hygrometrica*. That would simplify sequencing but necessitate testing its activity in *Physcomitrium*.