SUPPLEMENTARY INFORMATION

Twin-primer non-enzymatic DNA assembly: an efficient and accurate multi-part DNA assembly method

Jing Liang[1,§], Zihe Liu[1,§], Xi Zhi Low[3], Ee Lui Ang[1,*] & Huimin Zhao[1,2,*]

[1] Metabolic Engineering Research Laboratory, Science and Engineering Institutes, Agency for Science, Technology and Research, Singapore.
[2] Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.
[3] NUS High School of Mathematics and Science, Singapore.

* To whom correspondence should be addressed. Tel:+1 217 333 2631; Fax:+1 217 333 5052; Email: zhao5@illinois.edu. Correspondence may also be addressed to E.L.A. (email: angel@merl.a-star.edu.sg)

§ J.L. and Z.L. contributed equally.

Table S1. 9+1 white colonies gene sequencing results summary. Total bases sequenced = 7130, junction bases = 697.

Plasmid ID	Base number	Gene	DNA change	Amino acid change	Location of mutation
13	1899	CrtE	A ins	frameshift after 7	non-junction, coding
13	2004	CrtE	G to A	A 43 T	non-junction, coding
1	2840	CrtB	C del	frameshift after 6	junction, coding
14	3305	CrtB	T to C	C 162 R	junction, coding
9	3829	CrtZ	C del	frameshift after 167	non-junction, coding
13	4847	CrtY	G to A	A 222 V	non-junction, coding
9	4881	CrtY	G del	frameshift after 211	junction, coding
10	4881	CrtY	G del	frameshift after 211	junction, coding
13	5610	CrtI	T to C	H 468 P	non-junction, coding
2	6246	CrtI	GC del	frameshift after 256	junction, coding

Figure S1. Effect of *Dpn*I digestion and choice of competent cells. Plotted values are the average of two experiments and error bars represent the standard deviation. (A) Efficiency and fidelity of reaction with and without DpnI. Columns represent efficiency, pink line plot represents fidelity. (B) Transformation efficiency of 7+1 and 9+1 zeaxanthin pathway assemblies into two different competent cells. (C) Assembly fidelity of 7+1 and 9+1 zeaxanthin pathway assemblies into two different competent cells.

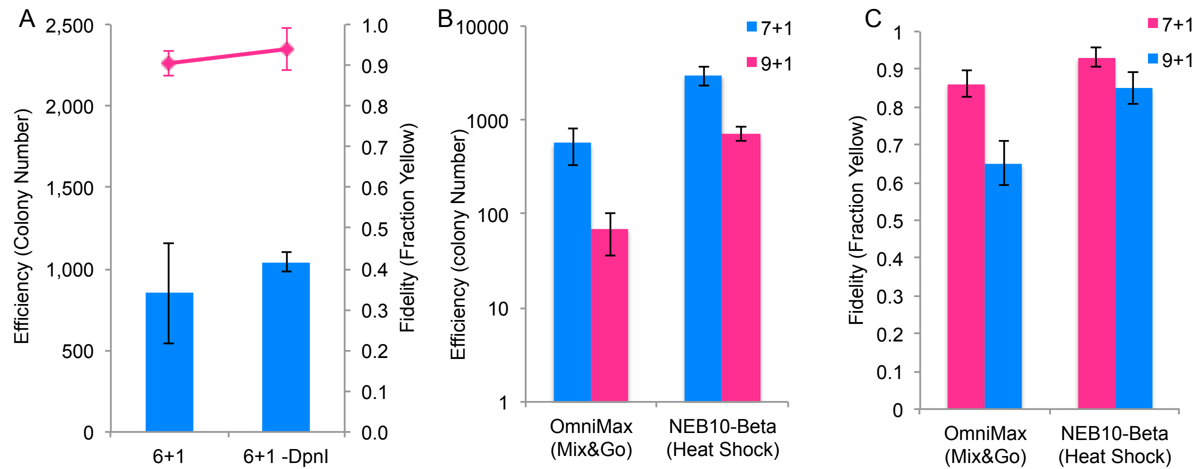
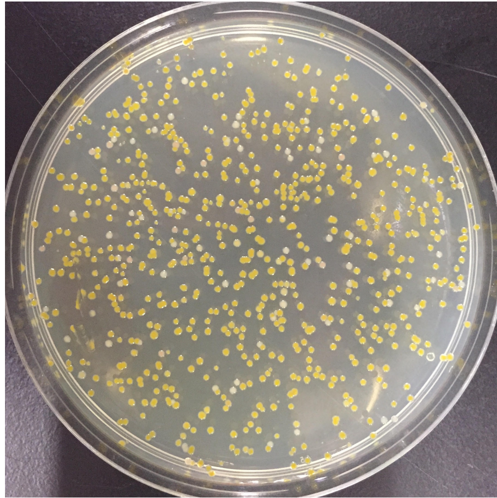


Figure S2. Restriction digest verification of yellow colonies from the zeaxanthin pathway TPA assembly. (A) Yellow and white colonies as seen on a 7+1 assembly plate. (B) Restriction digestion of the plasmid from 14 yellow colonies all showed the correct pattern.

A



B

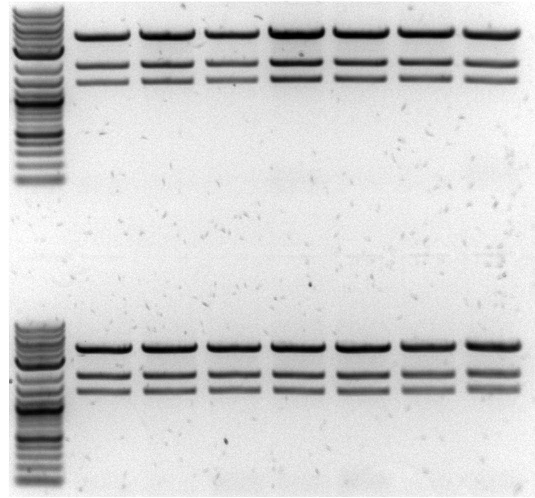


Figure S3. Vector files of all plasmids used in this study. (A) The 7 kb plasmid pAmp-EC-Zeax harbouring the zeaxanthin pathway. (B) The 16 kb plasmid pSSB100 harbouring the (R,R)-2,3-butanediol (BDO) pathway and GFP (1). (C) The 21 kb plasmid pSSB100-Zeax harbouring the zeaxanthin pathway as well as the BDO pathway and GFP (1). (D) The 31 kb plasmid pSSB-144 harbouring a butanol pathway (2). (E) The 6kb plasmid pAmp-BLUE3 that had two high GC homology arms (66% and 71% respectively) for genome integration in *Streptomyces.*

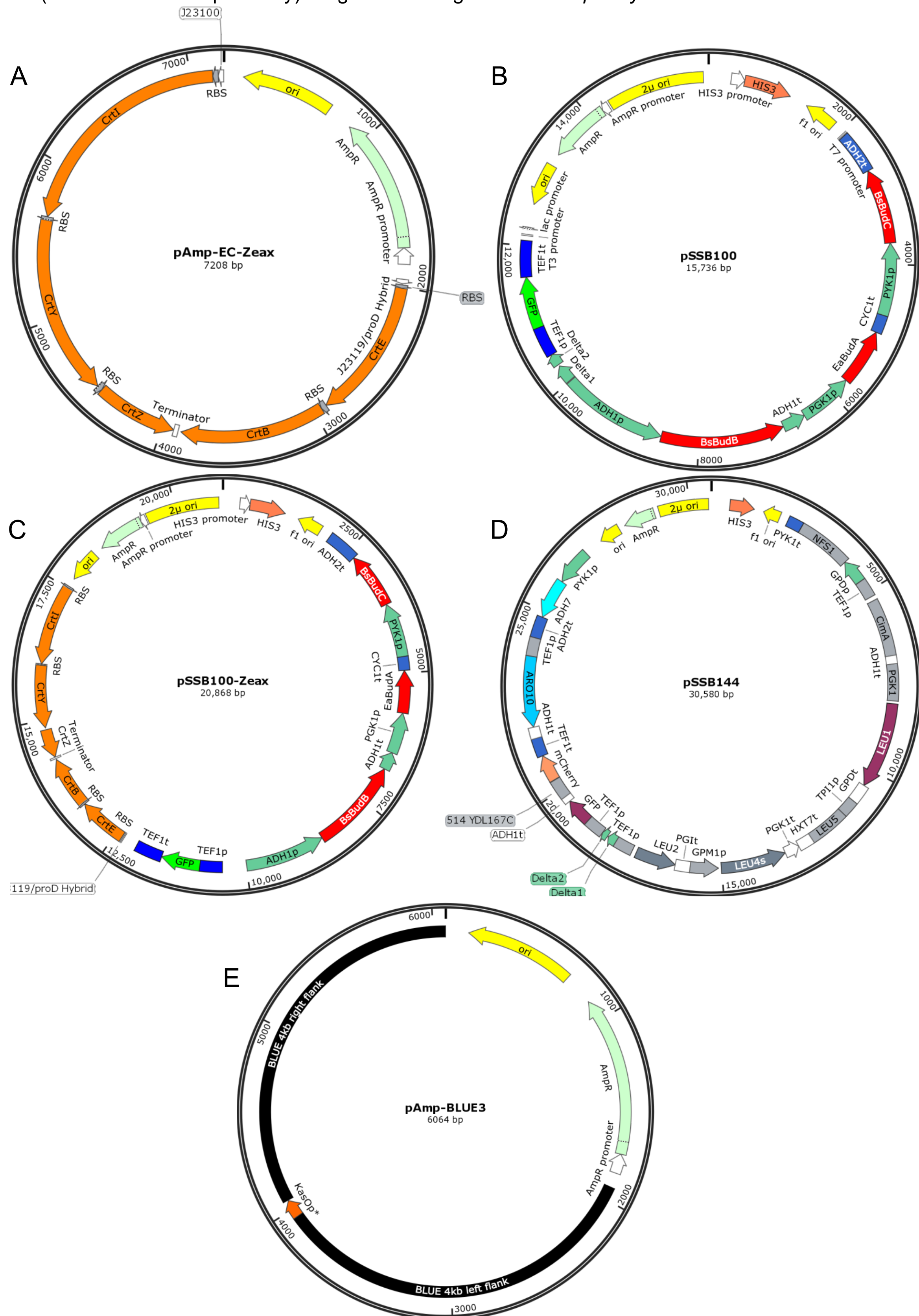


Figure S4. Sequence verification of TPA junctions for 21 kb and 31 kb assemblies. In each alignment, a double-stranded reference sequence is shown with the 4 TPA assembly primers and any features located close to the junction. Below the reference is a series of aligned single-stranded sequences. The bolded sequence is the reference whereas the none-bolded ones are the sequencing reads from different clones. (A1-5) 5 TPA junctions in the 21 kb assembly. Junction A2's differences are caused by a template-primer mismatch, they are synonymous and are not counted as errors. (B1-5) 5 TPA junctions in the 31 kb assembly.

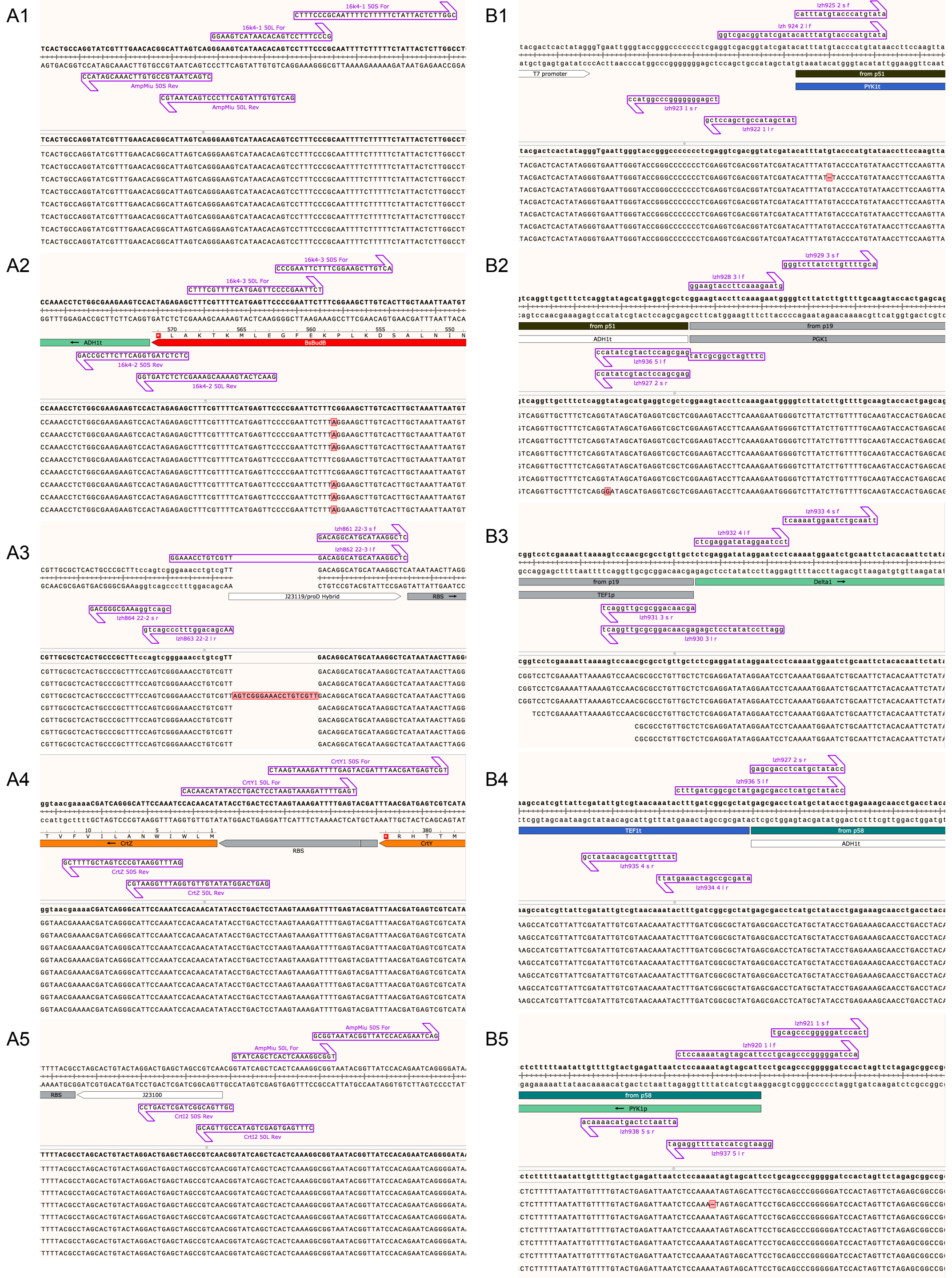


Figure S5. Sequence verification of TPA and GA junctions for 16 kb assemblies. (A1-5) 5 TPA junctions in the 16 kb assembly. Junction A3 has a primer-template mismatch. (B1-5) 5 GA junctions in the 16 kb assembly. Junction B3 has a primer-template mismatch.

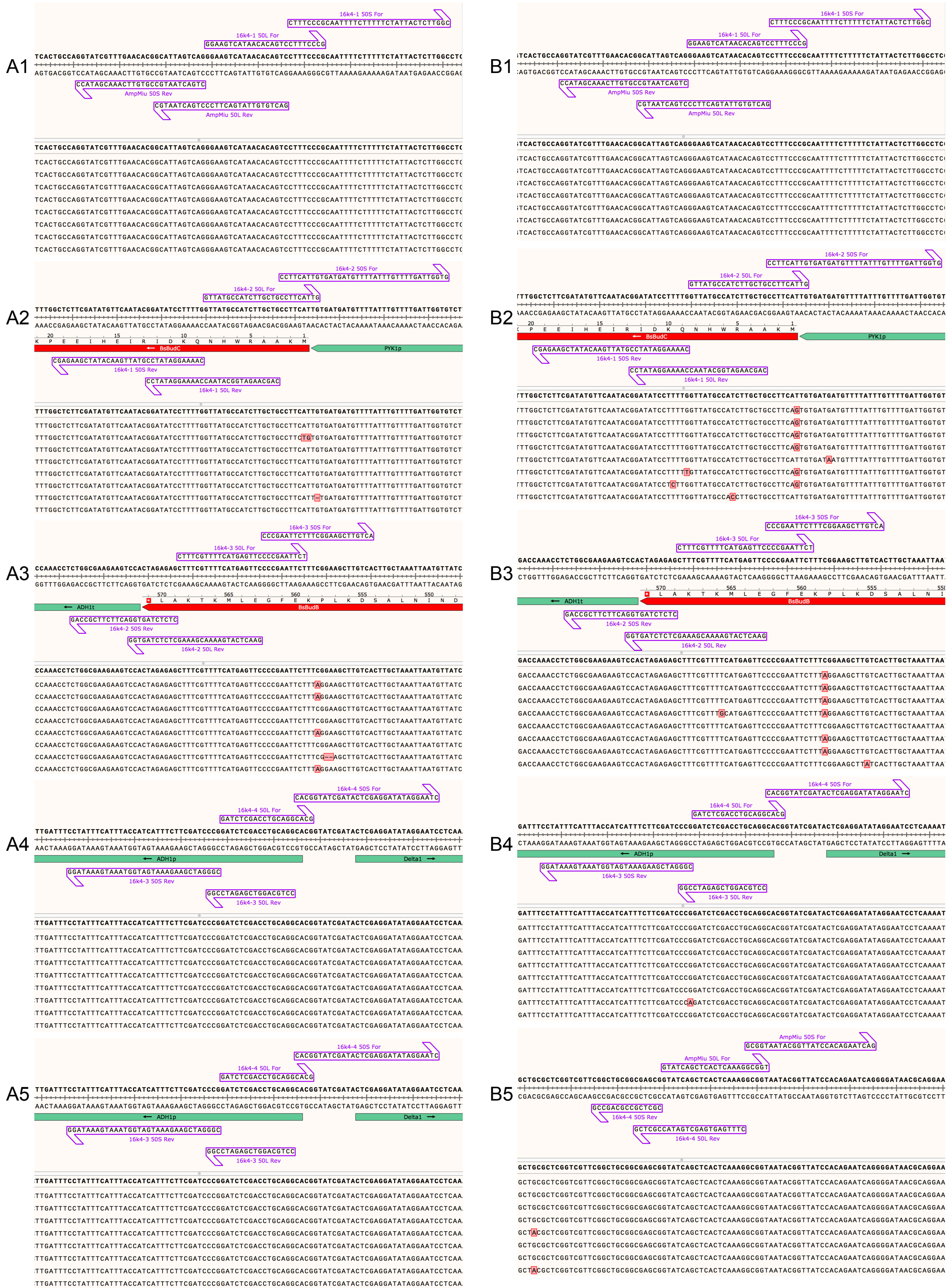


Figure S6. Restriction digest verification of TPA and GA assemblies. Wrong patterns are marked with a pink X. (A) TPA 21 kb assemblies. (B) TPA 31 kb assemblies. (C) TPA 16 kb assemblies §. (D) GA 16 kb assemblies §. (E) TPA GC assemblies §. (F) GA GC assemblies §.

§: Top panel used a different set of restriction enzymes.

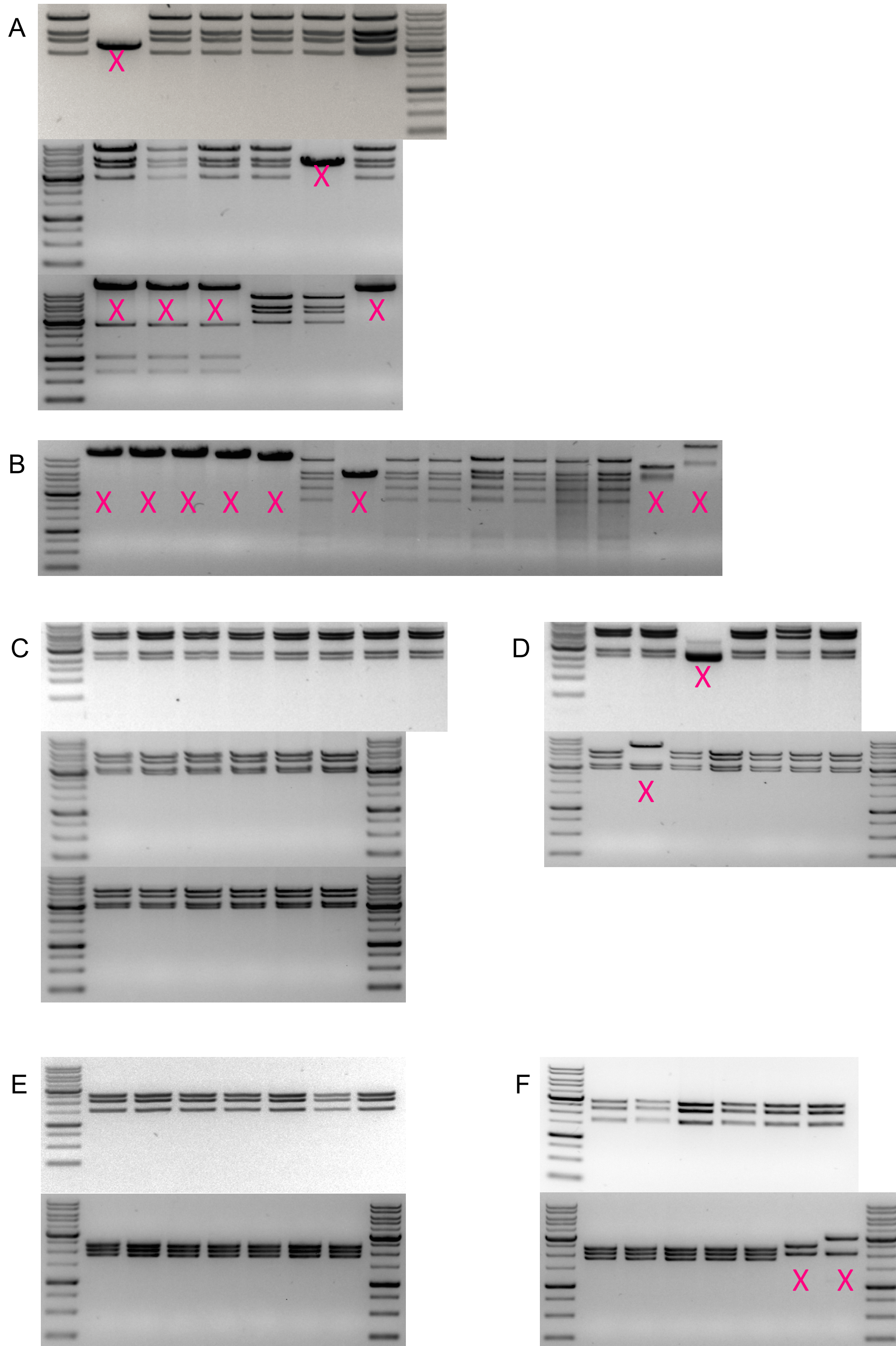


Figure S7. Effect of DNA Polymerase on fidelity. Below is a plate from the 4+1 zeaxanthin pathway TPA assembly showing nearly 100% yellow colonies.

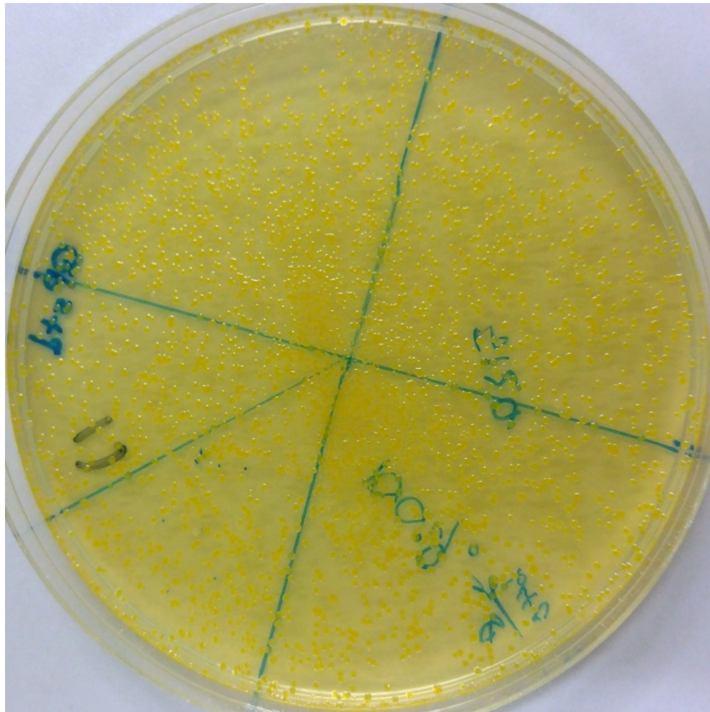
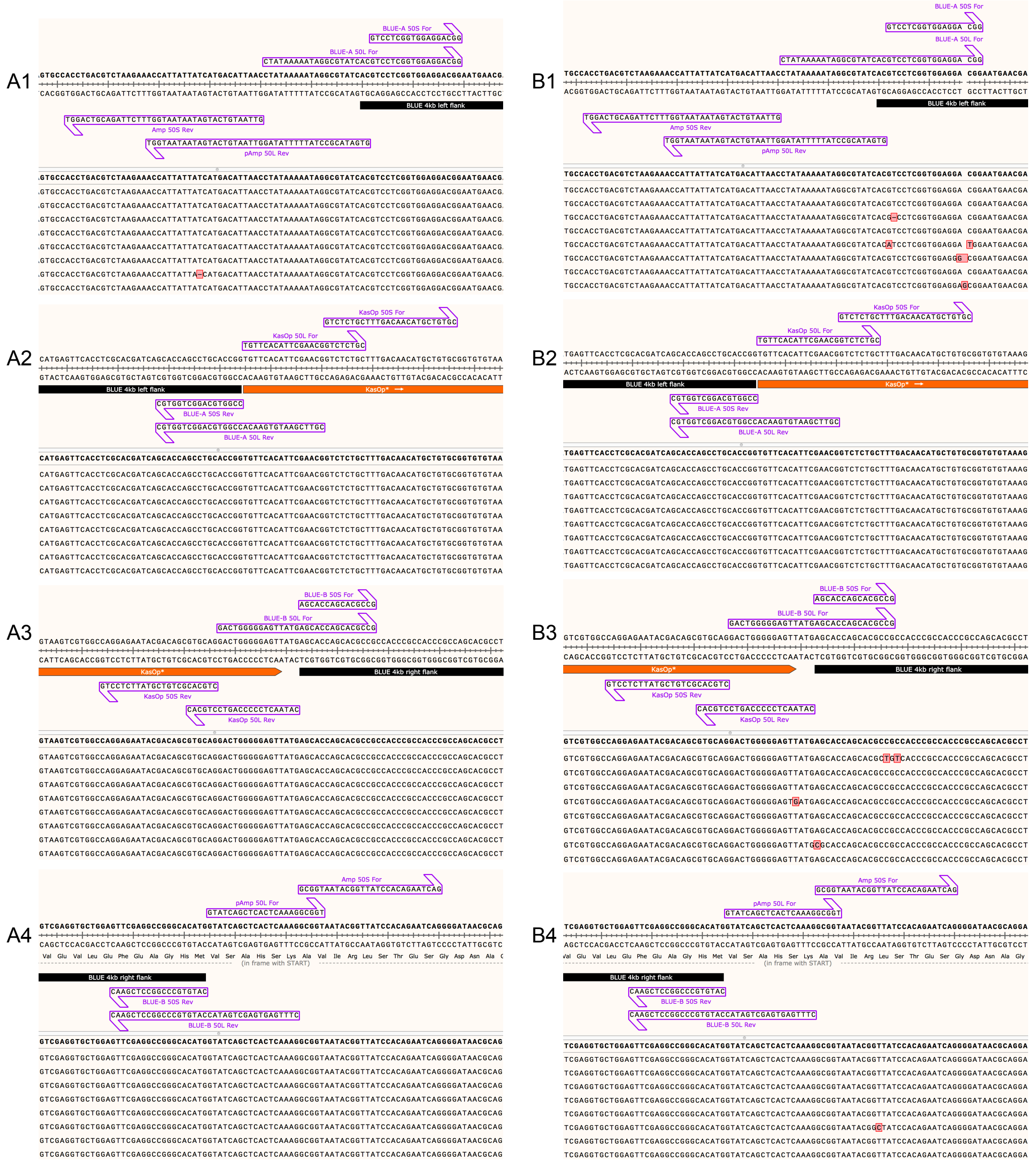
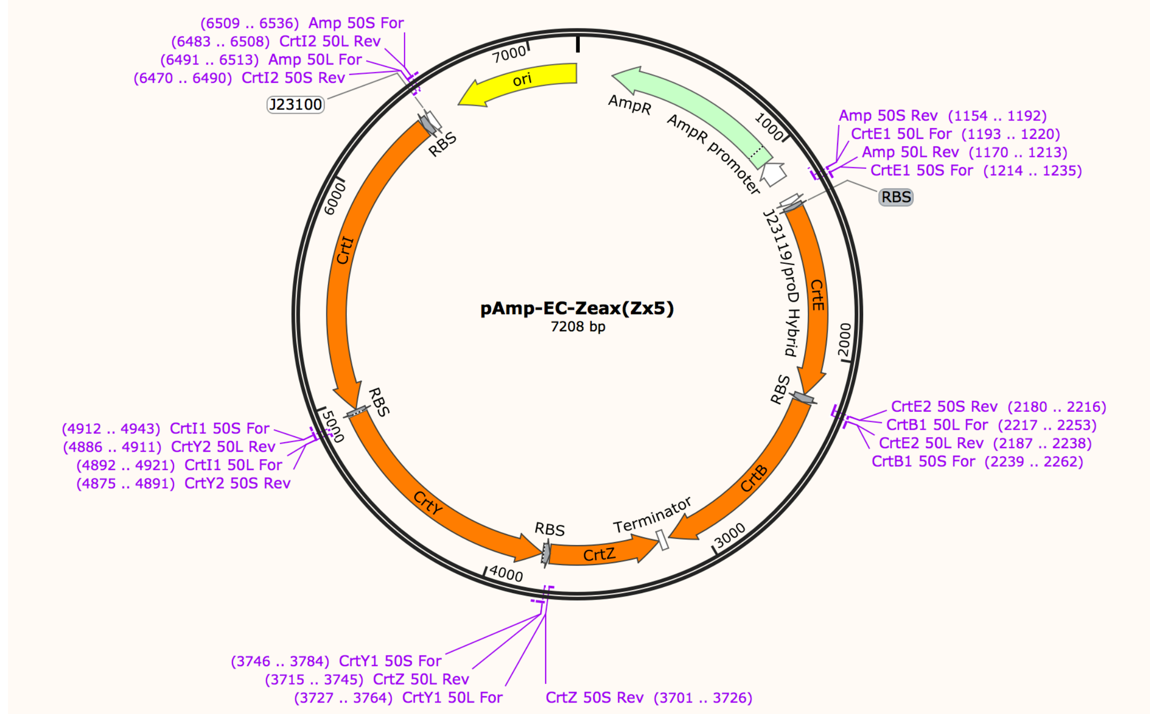


Figure S8. Sequence verification of TPA and GA junctions for GC assemblies. (A1-4) 4 TPA junctions in the GC assembly. (B1-4) 4 GA junctions in the GC assembly.



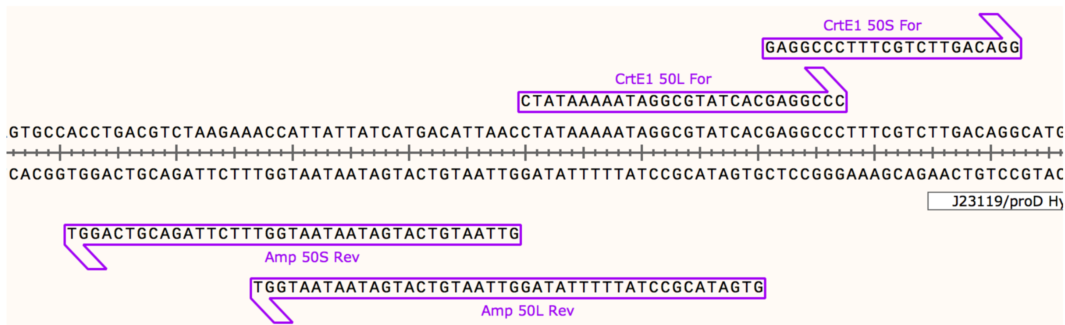
Supplementary Protocol

Below is a detail TPA protocol using the 5-fragment zeaxanthin pathway assembly as an example.

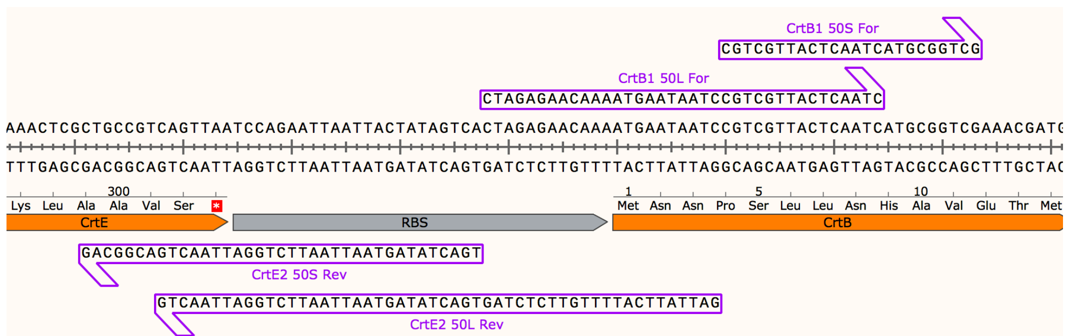


1. Primer Design
   Select the junction overlaps and design L and S primers around it. The five junctions are shown below. For this example, all primers fully anneal to the template because the final product is the template. For new constructions, L primers should have long enough template-annealing regions as recommended by the DNA polymerase's manufacturer, as well as junction regions that enable the assembly. Refer to Figure S4 B1-5 for examples of new construction primers.

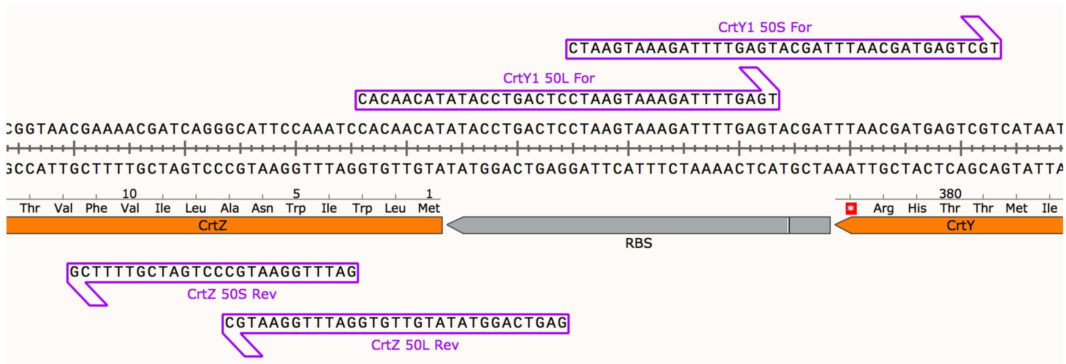
   Junction T_M is 50ºC for this assembly.



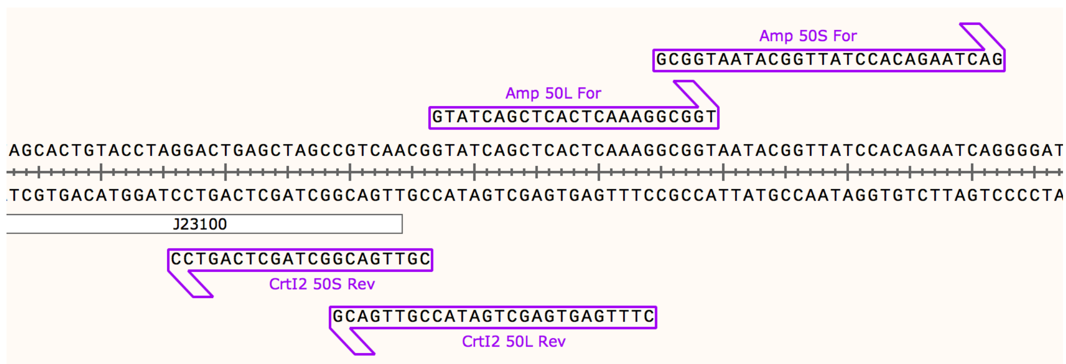
Junction 1



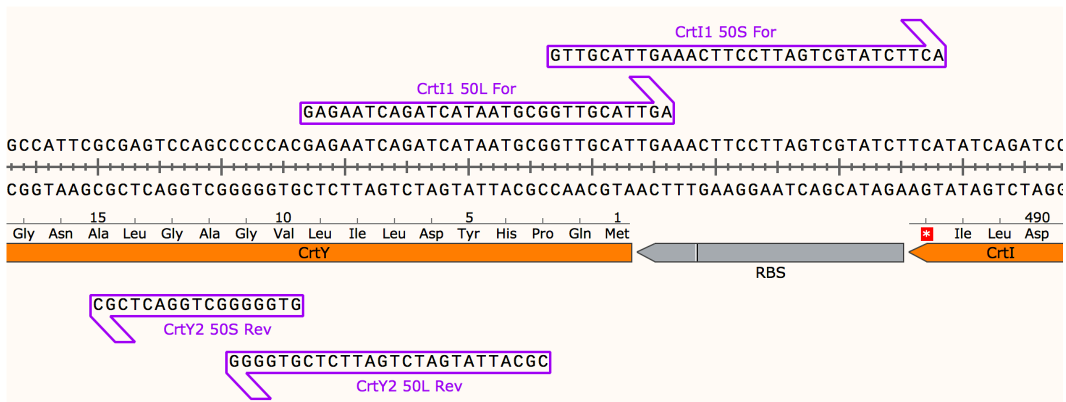
Junction 2



Junction 3



Junction 4



Junction 5

Well	Name	Sequence
A1	Amp 50L For	GTATCAGCTCACTCAAAGGCGGT
A2	Amp 50S For	GCGGTAATACGGTTATCCACAGAATCAG
A3	CrtE1 50L For	CTATAAAAATAGGCGTATCACGAGGCCC
A4	CrtE1 50S For	GAGGCCCTTTCGTCTTGACAGG
A7	CrtB1 50L For	CTAGAGAACAAAATGAATAATCCGTCGTTACTCAATC
A8	CrtB1 50S For	CGTCGTTACTCAATCATGCGGTCG
B1	Amp 50L Rev	GTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGT
B2	Amp 50S Rev	GTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGT
B5	CrtE2 50L Rev	GATTATTCATTTTGTTCTCTAGTGACTATAGTAATTAATTCTGGATTAACTG
B6	CrtE2 50S Rev	TGACTATAGTAATTAATTCTGGATTAACTGACGGCAG
B11	CrtZ 50L Rev	GAGTCAGGTATATGTTGTGGATTTGGAATGC
B12	CrtZ 50S Rev	GATTTGGAATGCCCTGATCGTTTTCG
C1	CrtY1 50L For	CACAACATATACCTGACTCCTAAGTAAAGATTTTGAGT
C2	CrtY1 50S For	CTAAGTAAAGATTTTGAGTACGATTTAACGATGAGTCGT
C5	CrtI1 50L For	GAGAATCAGATCATAATGCGGTTGCATTGA
C6	CrtI1 50S For	GTTGCATTGAAACTTCCTTAGTCGTATCTTCA
D3	CrtY2 50L Rev	CGCATTATGATCTGATTCTCGTGGGG
D4	CrtY2 50S Rev	GTGGGGGCTGGACTCGC
D7	CrtI2 50L Rev	CTTTGAGTGAGCTGATACCGTTGACG
D8	CrtI2 50S Rev	CGTTGACGGCTAGCTCAGTCC

Summarized list of the primers used in this assembly.

2. PCR
   Set up 10 PCR reactions according to the list below. We have mostly used KOD Xtreme Hot Start DNA Polymerase (EMD Millipore), but Q5 DNA Polymerase (NEB) has been shown to work perfectly fine as long as the a good quality product band can be obtained. Do NOT use non-proofreading DNA polymerases such as *Taq* DNA polymerase.

   If a plasmid template is used for PCR, use just 1ng template per reaction. This will help to reduce transformation background.

ID	Frag	Primer 1	Primer 2	Template	Length
1	Amp 50L	Amp 50L For == A1	Amp 50S Rev == B2	pAmp-EC-Zeax	1910
2	Amp 50R	Amp 50S For == A2	Amp 50L Rev == B1	pAmp-EC-Zeax	1913
3	CrtE 50L	CrtE1 50L For == A3	CrtE2 50S Rev == B6	pAmp-EC-Zeax	1024
4	CrtE 50R	CrtE1 50S For == A4	CrtE2 50L Rev == B5	pAmp-EC-Zeax	1025
5	CrtBZ 50L	CrtB1 50L For == A7	CrtZ 50S Rev == B12	pAmp-EC-Zeax	1510
6	CrtBZ 50R	CrtB1 50S For == A8	CrtZ 50L Rev == B11	pAmp-EC-Zeax	1507
7	CrtY 50L	CrtY1 50L For == C1	CrtY2 50S Rev == D4	pAmp-EC-Zeax	1165
8	CrtY 50R	CrtY1 50S For == C2	CrtY2 50L Rev == D3	pAmp-EC-Zeax	1166
9	CrtI 50L	CrtI1 50L For == C5	CrtI2 50S Rev == D8	pAmp-EC-Zeax	1599
10	CrtI 50R	CrtI1 50S For == C6	CrtI2 50L Rev == D7	pAmp-EC-Zeax	1597

3. PCR cleanup
   Gel-purify the PCR products. If PCR has a single clean band, PCR purification is also acceptable. Quantify the concentration of the eluted PCR products using Nanodrop. If the 260:230 ratio is below 1.6, re-purification using Zymo Research's DNA Cleanup kit is recommended.

4. Mix PCR products
   Use a spreadsheet to calculate the approximate molecular weight (MW) of the fragments using MW = length*607.4. Calculate the volume of PCR product needed to make 400 fmol, and the volume of water needed to make 15 μL.

ID	Frag	Length	Approx MW	10fmol in ng	Conc (ng/uL)	400fmol in uL	H2O to 15uL	Premix
1	Amp 50L	1910	1160134	11.60134	154.5	3.00		
2	Amp 50R	1913	1161956.2	11.619562	163.1	2.85	9.15	K1
3	CrtE 50L	1024	621977.6	6.219776	157.5	1.58		
4	CrtE 50R	1025	622585	6.22585	150.1	1.66	11.76	K2
5	CrtBZ 50L	1510	917174	9.17174	172.4	2.13		
6	CrtBZ 50R	1507	915351.8	9.153518	159	2.30	10.57	K3
7	CrtY 50L	1165	707621	7.07621	155.5	1.82		
8	CrtY 50R	1166	708228.4	7.082284	151.1	1.87	11.30	K4
9	CrtI 50L	1599	971232.6	9.712326	181	2.15		
10	CrtI 50R	1597	970017.8	9.700178	175.7	2.21	10.65	K5

5. Add the reaction master mix
   Prepare a master mix consisting of 10 μL 10X CutSmart buffer, 2.5 μL *Dpn*I, and 12.5 μL water. The addition of *Dpn*I is recommended if plasmids have been used as PCR templates. Add 5 μL of the master mix to each of the 15 μL premixes to make 40 fmol/μL fragments. Mix well.

6. For 2-Step protocol, re-anneal these premixes on a thermocycler using the following profile: 37 °C for 30 min (for *Dpn*I digest), 98 °C for 2 min, 85 °C for 2 min (at a ramping rate of 0.1 °C/s), 75 °C for 2 min (0.1 °C/s), 65 °C for 2 min (0.1 °C/s), 55 °C for 2 min (0.1 °C/s), and 8 °C hold (0.1 °C /s). For 1-Step protocol, skip to 7.

7. Prepare the hybridization reaction
   Prepare 1X CutSmart buffer. Then set up 1 assembly reaction and 2 negative control reactions using by putting together the following:

   Assembly: 1 μL each of K1, K2, K3, K4, K5, and 5 μL 1X CutSmart
   Negative 1: 1 μL K1 and 9 μL 1X CutSmart
   Negative 2: 1 μL each of K2, K3, K4, K5, and 6 μL 1X CutSmart

8. Hybridize using either the 1-Step protocol or the 2-Step protocol.

   1-Step: 37 °C for 30 min (*Dpn*I digest), 98 °C for 2 min, 85 °C for 2 min (0.1 °C/s), 75 °C for 2 min (0.1 °C /s), 65 °C for 2 min (0.1 °C/s), 55.5 °C (T_M+5.5) for 1 to 2 hour (0.1 °C/s), and 8 °C hold (0.1 °C/s).

   2-Step: 65 °C for 10 seconds, 55.5 °C for 2 hours, and 8 °C hold (0.1 °C/s).

9. Transform as soon as possible. Avoid overnight storage as efficiency may suffer. TPA is compatible with either Mix & Go transformation or heat shock chemical transformation. Do not use electroporation.

Supplementary Discussion

F(Inactive) = Fraction of inactive zeaxanthin pathway
P(JE) = Probability of junction error
E(PE) = Expected number of PCR error in assembly
$F(JC)_i$ = Fraction of coding bases within junction i
F(JC)j = Fraction of coding bases within junction j
$F(JN)_i = 1 – F(JC)_i$
$F(JC)_j = 1 – F(JN)_j$
$P(EE_{cd})$ = Probability of inactivation if an error occurs within coding region
$P(EE_{nc})$ = Probability of inactivation if an error occurs within non-coding region
F(PC) = Fraction of PCR amplified bases within coding region
F(PN) = 1 – F(PC)

$$F(Inactive) = E(PE)[F(PC)P(EE_{cd}) + F(PN)P(EE_{nc})] + \sum P(JE)[F(JC)_i P(EE_{cd}) + F(JN)_i P(EE_{nc})] - \sum P(JE)^2 [F(JC)_i P(EE_{cd}) + F(JN)_i P(EE_{nc})]\left[F(JC)_j P(EE_{cd}) + F(JN)_j P(EE_{nc})\right] - \cdots$$

Assume that non-coding region error only has a negligible chance of inactivating the entire pathway.
Assume that proportion of pathways containing 2 or more inactivating junction error is negligible.

$$F(Inactive) \approx E(PE)F(PC)P(EE_{cd}) + \sum P(JE)F(JC)_i P(EE_{cd})$$

From KOD 4+1 zeaxanthin and Q5 4+1 zeaxanthin assembly, junction terms can be eliminated since they share the same junctions.

$$F(Inactive)_{KOD} - F(Inactive)_{Q5} \approx [E(PE)_{KOD} - E(PE)_{Q5}]F(PC)P(EE_{cd})$$

Assume 30 effective PCR cycles, 7200bp amplified bases.
Assume Q5 has an error rate in the range of Phusion (3), and KOD has an error rate in the range of Pfu (4,5).
We have the following estimates(6):

E(PE)Q5 = 0.095
E(PE)KOD = 0.60
F(PC) = 5000/7200

From our experimental data:

F(Inactive)KOD ≈ 0.064
F(Inactive)Q5 ≈ 0.005

$$0.059 \approx [0.6 - 0.095]\left(\frac{5000}{7200}\right) P(EE_{cd})$$

We obtain an empirical estimate of $P(EE_{cd})$.

$P(EE_{cd}) = 0.17$

$$F(Inactive) \approx E(PE)F(PC)(0.17) + \sum P(JE)F(JC)_i(0.17)$$

From all our junction sequencing data:
Total detected junction error/Total sequenced junctions = 10/145

P(JE) = 0.069

$$F(Inactive) \approx E(PE)F(PC)(0.17) + \sum (0.069)F(JC)_i(0.17)$$

From an analysis of our 10 junctions used to assembly the zeaxanthin pathway plasmid, we calculated the fraction of coding sequence within each junction region.

F(JC)1	F(JC)2	F(JC)3	F(JC)4	F(JC)5	F(JC)6	F(JC)7	F(JC)8	F(JC)9	F(JC)10
0	1	0.583	1	1	0.578	1	0.681	1	0

For KOD 4+1 assembly:

$$F(Inactive) \approx (0.6)\left(\frac{5000}{7200}\right)(0.17) + \sum (0.069)F(JC)_i(0.17)$$

$$F(Inactive) \approx (0.6)\left(\frac{5000}{7200}\right)(0.17) + (0.069)(0.17)(0 + 0.58 + 0 + 0.58 + 0.68)$$

$$F(Inactive) \approx 0.09$$

For Q5 4+1 assembly:

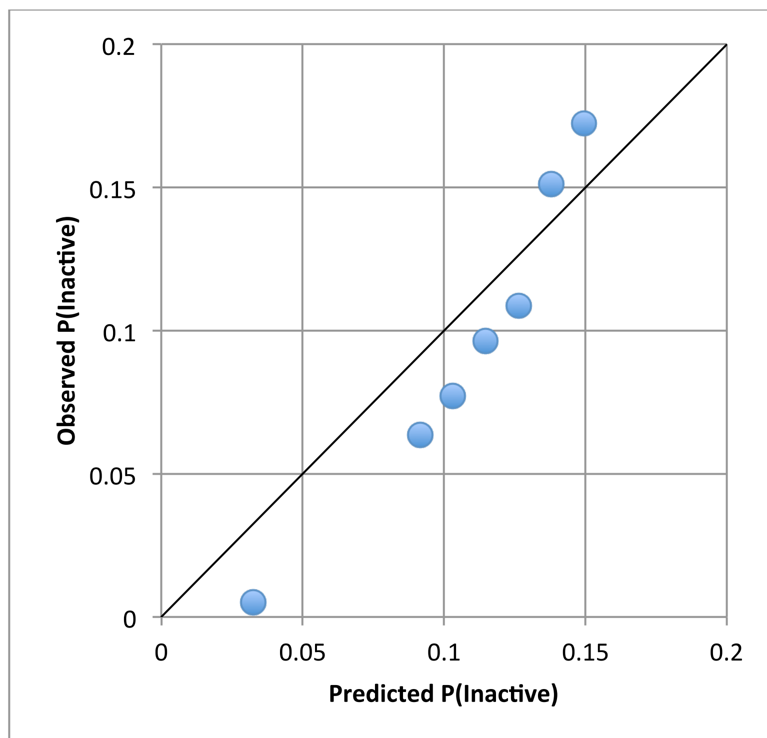
$$F(Inactive) \approx (0.095)\left(\frac{5000}{7200}\right)(0.17) + (0.069)(0.17)(0 + 0.58 + 0 + 0.58 + 0.68)$$

$$F(Inactive) \approx 0.03$$

For KOD 9+1 assembly:

$$F(Inactive) \approx 0.15$$

Validation:
The available observed data points from our various zeaxanthin pathway assemblies are compared to the predicted values. The data points from low to high are Q5 4+1, KOD 4+1, 5+1, 6+1, 7+1, 8+1, 9+1.



References

1. Shi, S., Liang, Y., Zhang, M.M., Ang, E.L. and Zhao, H. (2016) A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in *Saccharomyces cerevisiae*. *Metab. Eng.*, **33**, 19-27.
2. Shi, S., Si, T., Liu, Z., Zhang, H., Ang, E.L. and Zhao, H. (2016) Metabolic engineering of a synergistic pathway for n-butanol production in *Saccharomyces cerevisiae*. *Sci. Rep.*, **6**, 25675.
3. New England Biolabs. (2017). DNA Polymerase Selection Chart. *http://www.neb.sg/tools-and-resources/selection-charts/dna-polymerase-selection-chart*.
4. Merck. (2016). KOD Xtreme™ Hot Start DNA Polymerase. *https://www.merckmillipore.com/SG/en/product/KOD-Xtreme%E2%84%A2-Hot-Start-DNA-Polymerase,EMD_BIO-71975*.
5. Toyobo. (2016). High Fidelity & Efficient PCR Enzyme: KOD DNA polymerase. *http://www.toyobo-global.com/seihin/xr/lifescience/technology/001.html*.
6. Thermo Fisher Scientific. (2016). PCR Fidelity Calculator. *https://www.thermofisher.com/sg/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/pcr-fidelity-calculator.html*.