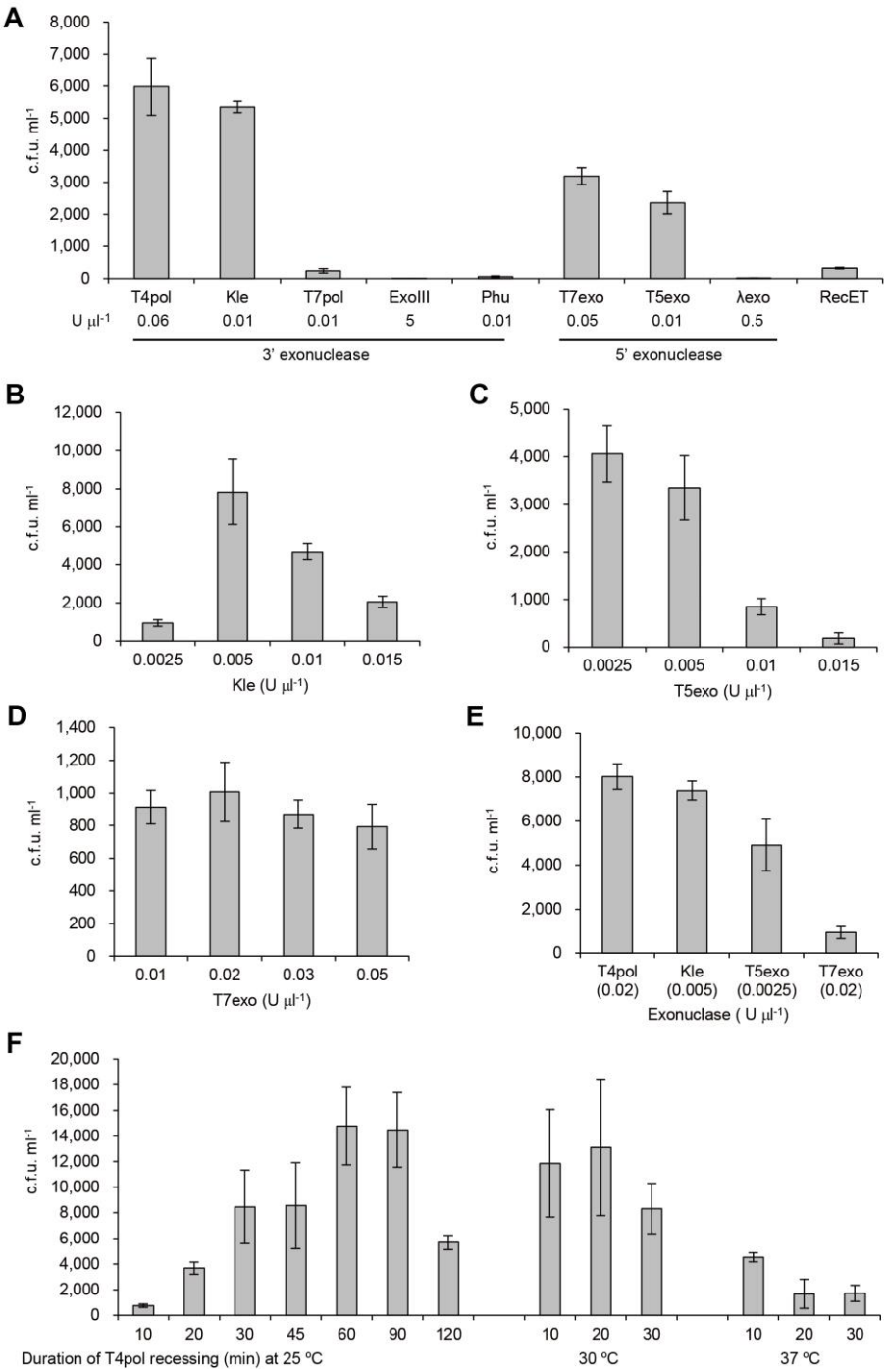
Supporting Information

ExoCET: Exonuclease *in vitro* assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes

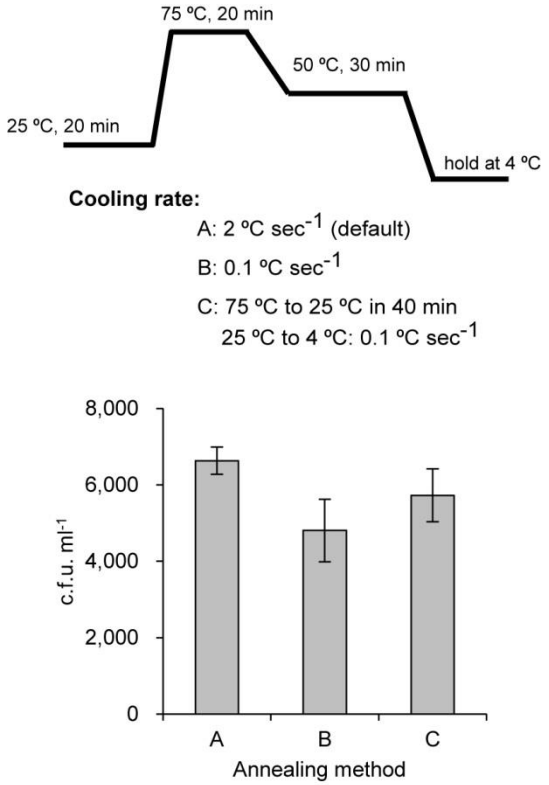
Hailong Wang[1], Zhen Li[1], Ruonan Jia[1], Jia Yin[1], Aiying Li[1], Liqiu Xia[3], Yulong Yin[3], Rolf Müller[1,4], Jun Fu[1,*], A. Francis Stewart[2,*] & Youming Zhang[1,*]

[1]Shandong University–Helmholtz Institute of Biotechnology, State Key Laboratory of Microbial Technology, School of Life Science, Shandong University, Shanda Nanlu 27, 250100 Jinan, People's Republic of China
[2]Genomics, Biotechnology Center, Technische Universität Dresden, Tatzberg 47-51, 01307 Dresden, Germany
[3]College of Life Science, Hunan Normal University, 410081 Changsha, People's Republic of China
[4]Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Pharmaceutical Biotechnology, Saarland University, Campus C2 3, 66123 Saarbrücken, Germany

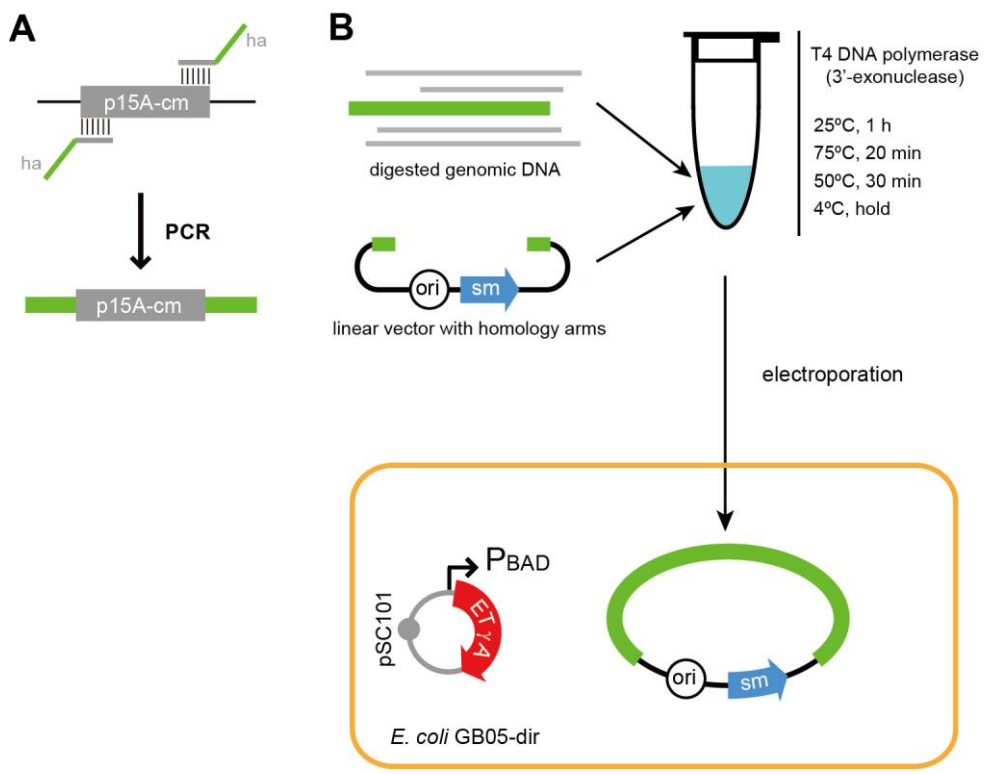
This file includes:



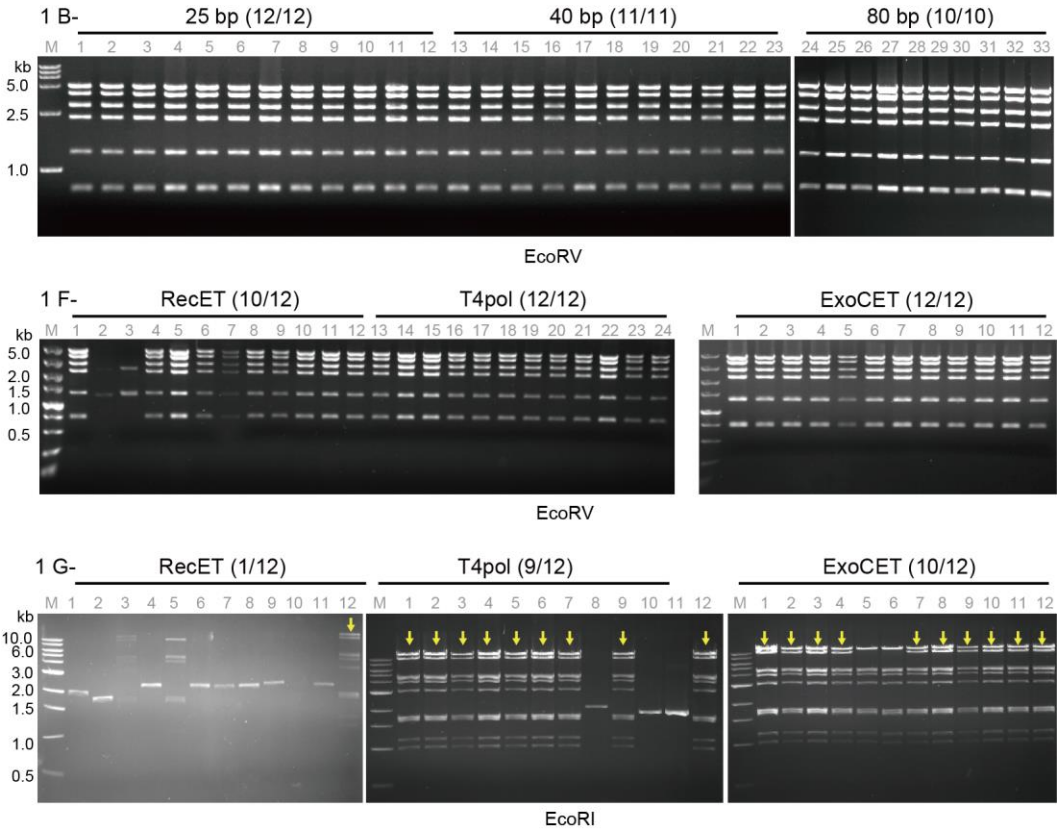
Supplementary Figure S1. Tests of different exonucleases for direct cloning of the 14 kb *lux* gene cluster. The p15A-cm vector and *P. phosphoreum* genomic DNA were treated with exonucleases before annealing and electroporation into arabinose induced *E. coli* GB05-dir. **(A)** Initial tests using the indicated exonucleases. **(B-D)** Optimization of the concentration of Kle, T5exo and T7exo. **(E)** Comparison of the cloning efficiency using the optimal amount of T4pol, Kle, T5exo and T7exo after 20 min of exonuclease digestion. **(F)** Optimization of the recessing temperature and duration of T4pol (0.02 U μl^{-1}). Error bars, s.d.; *n* = 3.



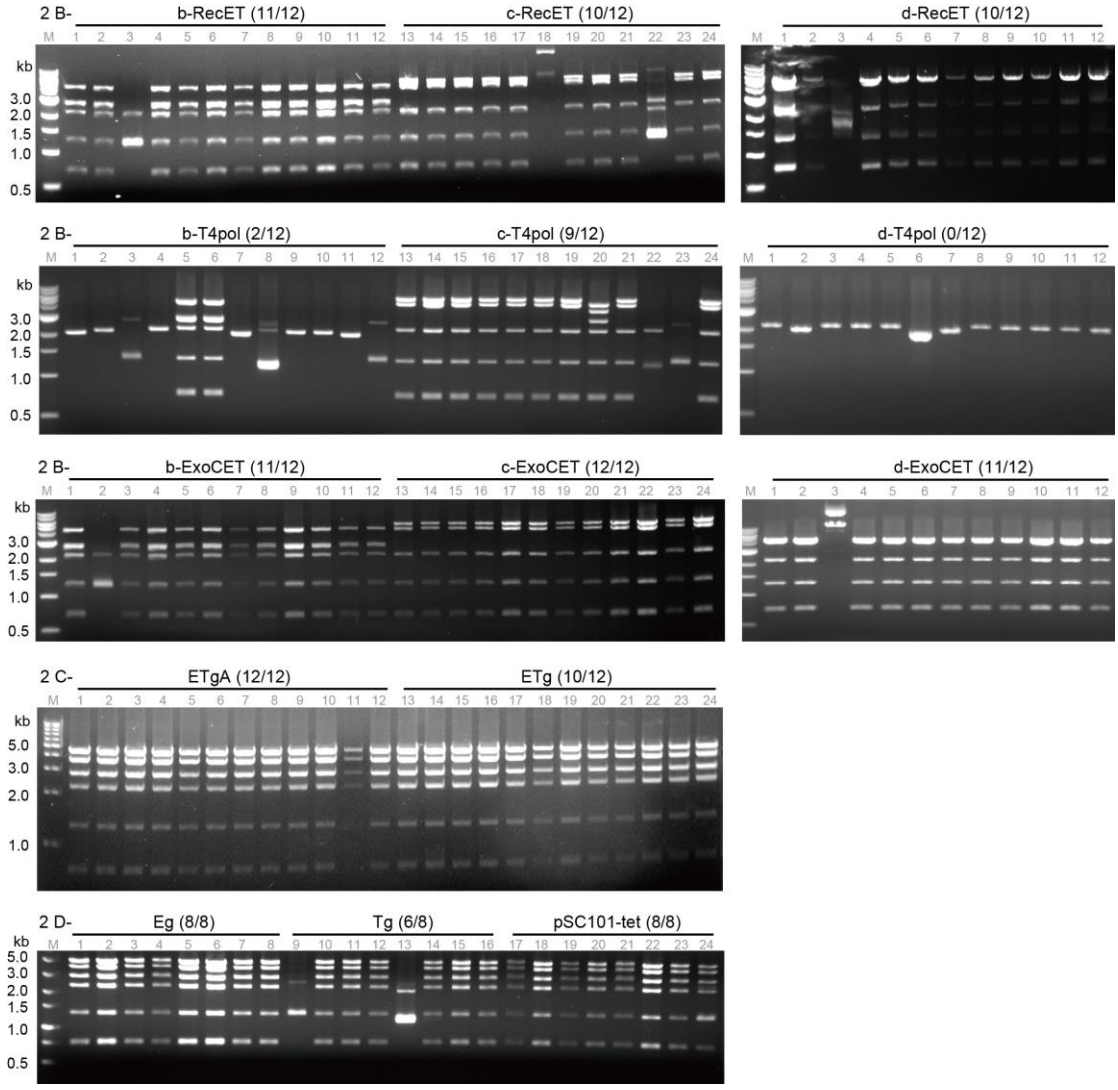
Supplementary Figure S2. Effects of cooling rate for annealing on direct cloning of the 14 kb *lux* gene cluster. Error bars, s.d.; *n* = 3. The p15A-cm vector and *P. phosphoreum* genomic DNA were treated with 0.02 U μl^{-1} T4pol and electroporated into arabinose induced *E. coli* GB05-dir after the different annealing protocols indicated as A, B or C.



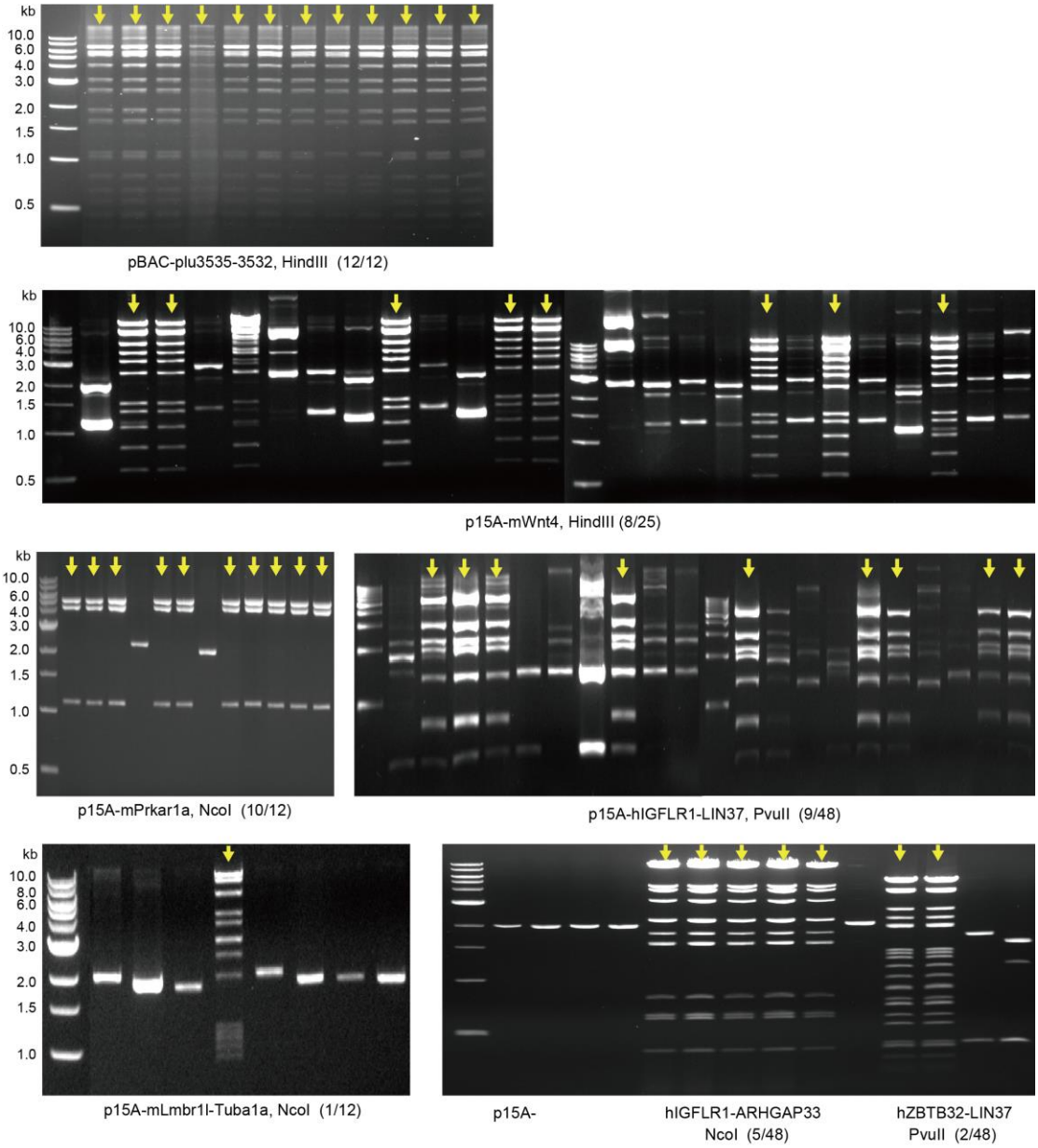
Supplementary Figure S3. Preparation of linear cloning vectors. (A) The p15A-cm vector was amplified by PCR using oligonucleotides including 80 nucleotides homology arms. **(B)** Standardized strategy for direct cloning from digested genomic DNA using ExoCET. The *in vitro* assembly products are electroporated into arabinose induced *E. coli* GB05-dir harboring pSC101-BAD-ETgA-tet. Correct recombinants are selected on LB plates containing appropriate antibiotics.



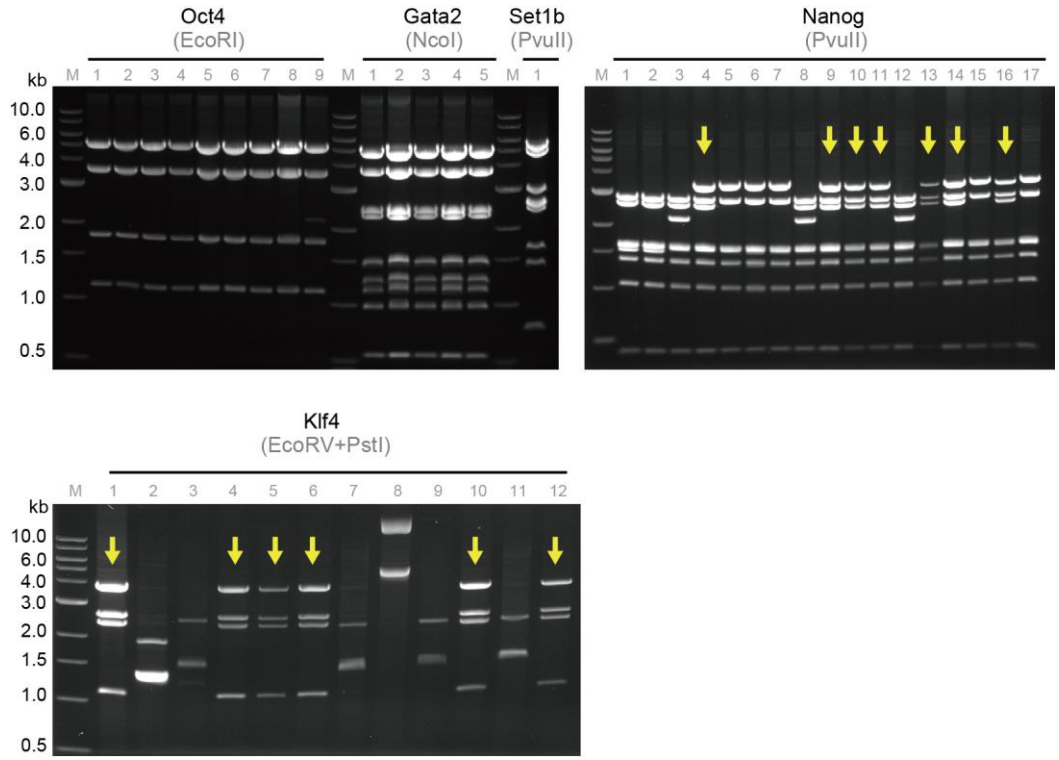
Supplementary Figure S4. Restriction analysis of clones obtained in Figure 1.



Supplementary Figure S5. EcoRV restriction analysis of clones obtained in Figure 2.

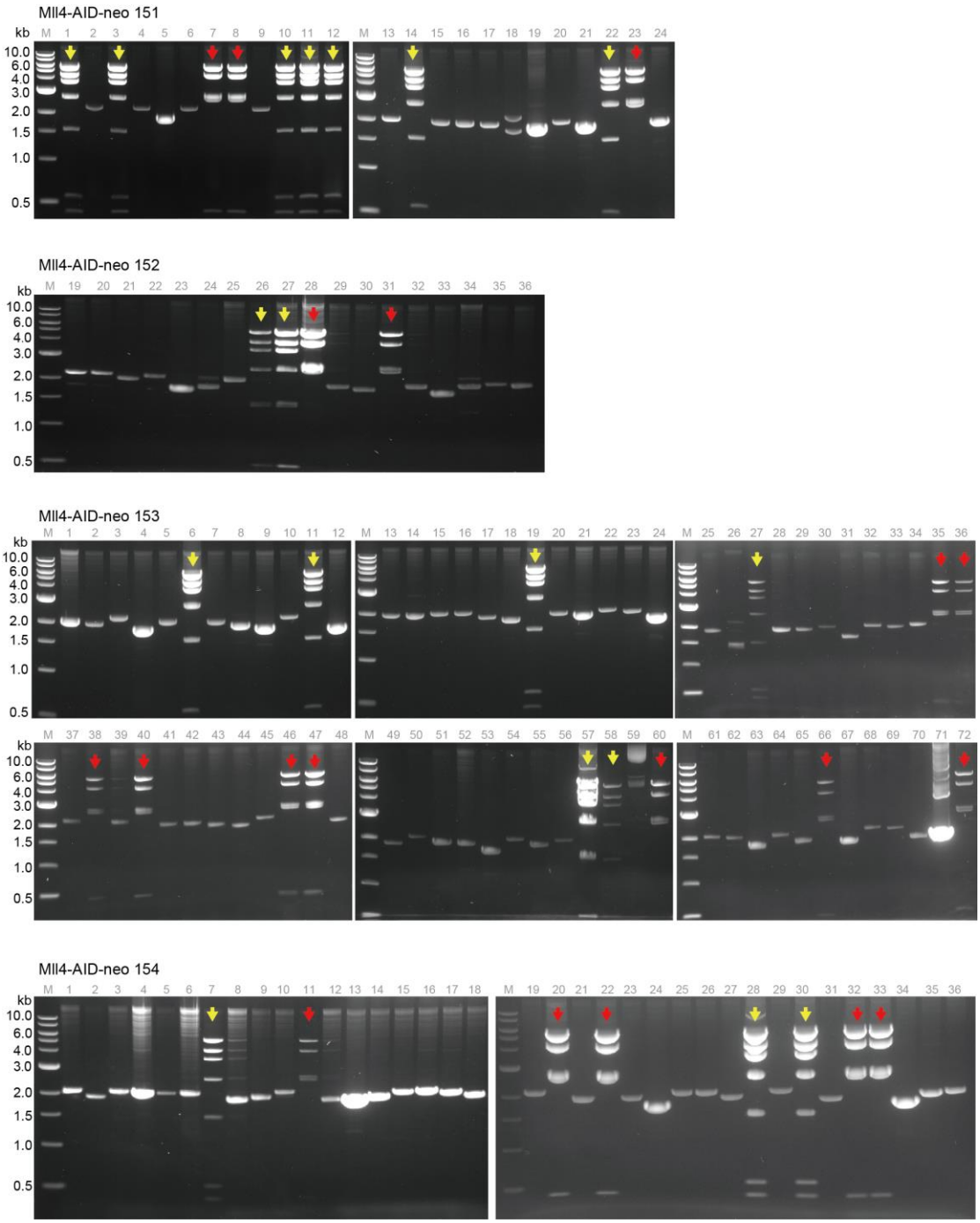


Supplementary Figure S6. Restriction analysis of clones obtained in Table 1. Correct clones are indicated with arrows.

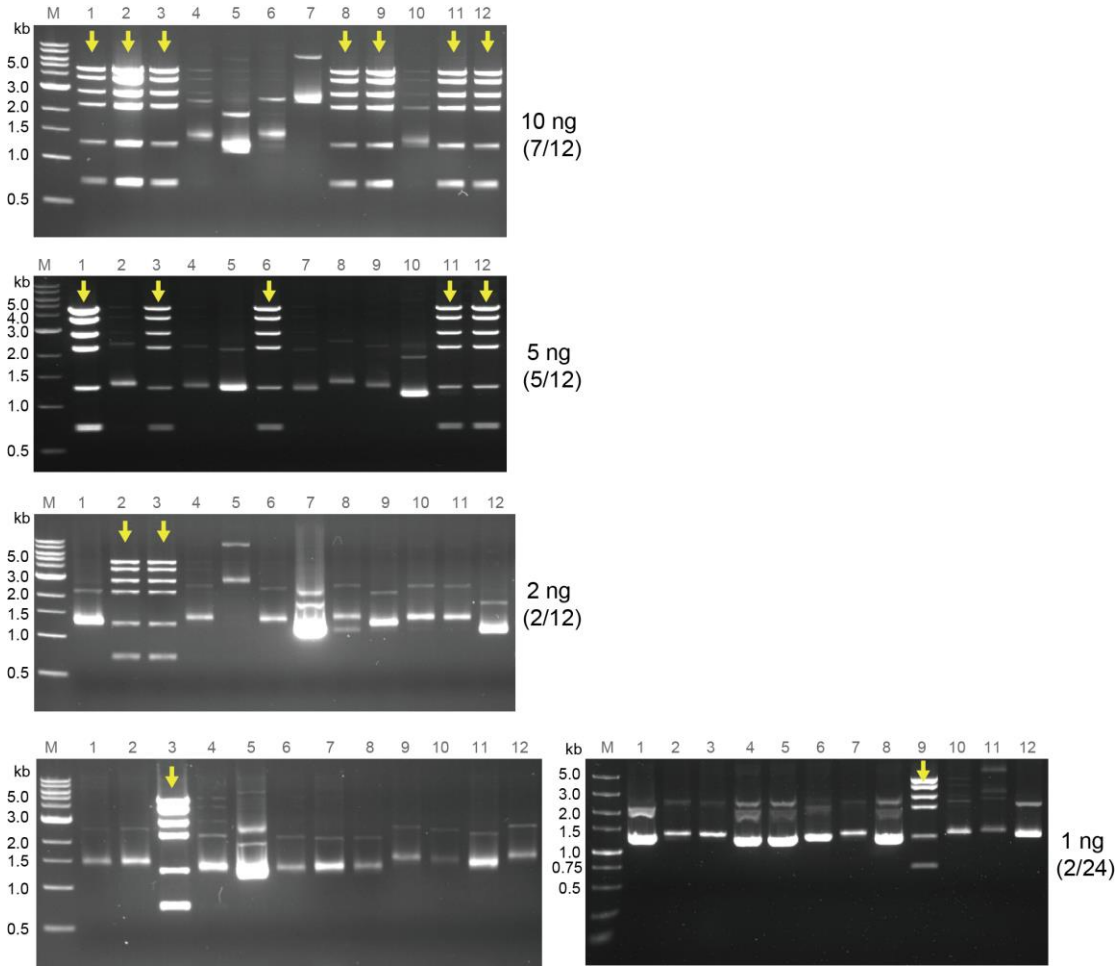


Supplementary Figure S6 continued. Restriction analysis of obtained clones in Table 1.

Correct clones are indicated with arrows (all Oct4 and Gata2 clones are correct).

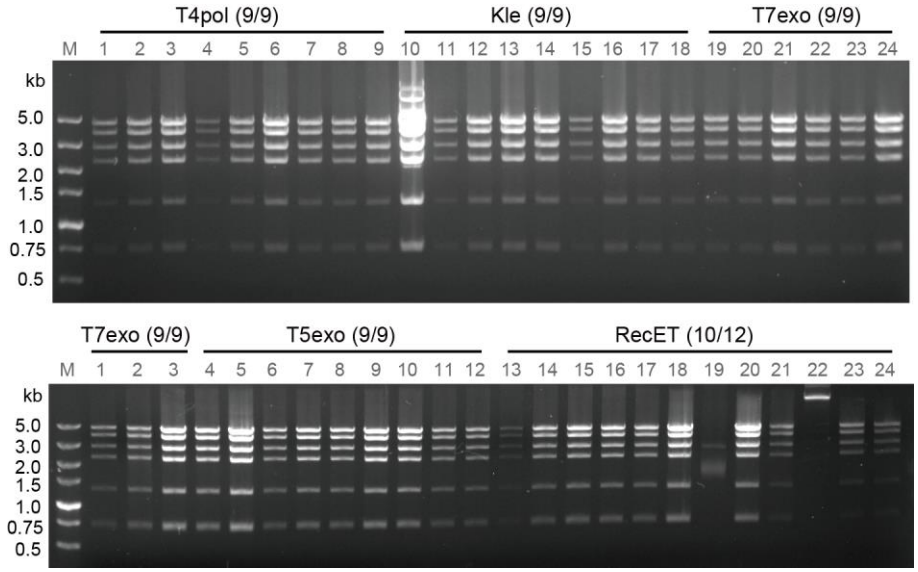


Supplementary Figure S6 continued. NcoI restriction analysis of clones obtained in Table 1. Clones containing targeted alleles are indicated with yellow arrows, and clones containing wild-type alleles are indicated with red arrows.

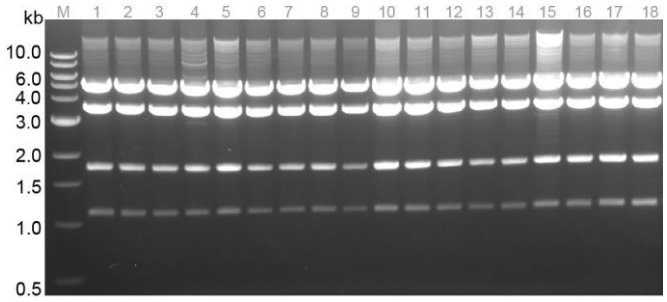


Supplementary Figure S7. EcoRV restriction analysis of clones obtained in Table 2.

Correct clones are indicated with arrows



Supplementary Figure S8. EcoRV restriction analysis of clones obtained in Supplementary Figure S1A.



Supplementary Figure S9. EcoRI restriction analysis of clones obtained in Supplementary Table S6.

Supplementary Note S1. The formula used to calculate the percentage of linear and circular products in the *in vitro* assembly reaction.

$$\text{linear } \% = \frac{\text{c. f. u.}_{ExoCET} - \text{c. f. u.}_{T4pol}}{\text{c. f. u.}_{ExoCET}} \times 100\%$$

$$\text{circular } \% = \frac{\text{c. f. u.}_{T4pol}}{\text{c. f. u.}_{ExoCET}} \times 100\%$$

In **Figure 1F**:

$$\text{linear } \% = \frac{32{,}500 - 4{,}880}{32{,}500} \times 100\% = 85\%$$

$$\text{circular } \% = \frac{4{,}880}{32{,}500} \times 100\% = 15\%$$

In **Figure 2B-a**:

$$\text{linear } \% = \frac{28{,}233 - 3{,}980}{28{,}233} \times 100\% = 86\%$$

$$\text{circular } \% = \frac{3{,}980}{28{,}233} \times 100\% = 14\%$$

Supplementary Table S1. Oligonucleotides for *in vitro* digestion of *S. albus* genomic DNA using Cas9-gRNA complexes

Name	**sequence (5'-3')**	
sal-gRNA-A1F	GACTGACACTGA**TAATACGACTCACTATA**GGCCAGTCTCCGTCCGGCCTgtttAagagctaTGCTGGAAACAG	
sal-gRNA-A2F	GACTGACACTGA**TAATACGACTCACTATA**GGCCTTCACCAGCCGCGACAgtttAagagctaTGCTGGAAACAG	
sal-gRNA-A3F	GACTGACACTGA**TAATACGACTCACTATA**GGGAGCGCGAACCCGGCTCCgtttAagagctaTGCTGGAAACAG	
sal-gRNA-A4F	GACTGACACTGA**TAATACGACTCACTATA**GGGGCTGGCTCTGAGCCGTTgtttAagagctaTGCTGGAAACAG	
sal-gRNA-B5F	GACTGACACTGA**TAATACGACTCACTATA**GGGGCACCGTGGGGTCGATAgtttAagagctaTGCTGGAAACAG	
sal-gRNA-B6F	GACTGACACTGA**TAATACGACTCACTATA**GGCGACATCGAGGACGACTTgtttAagagctaTGCTGGAAACAG	
sal-gRNA-B7F	GACTGACACTGA**TAATACGACTCACTATA**GGAGTTCGCGGGCGAGGCGGgtttAagagctaTGCTGGAAACAG	
sal-gRNA-B8F	GACTGACACTGA**TAATACGACTCACTATA**GGTCAGATATCCGGCGAGGGgtttAagagctaTGCTGGAAACAG	
gRNA-R	AAAAGCACCGACTCGGTGCCAC	
salTARGET-A-1	GTGATGCTGTCCGGGATCACCGAG	1602 bp
salTARGET-A-2	GCCGAGGACATCGACACGGTGATG	
salTARGET-B-1	GTGCCGCACTCATCCCCCAACTG	1500 bp
salTARGET-B-2	ACTCGGTGAAGGGGAACTGCATG	

The T7 promoter sequence is indicated in red. The CRISPR guide sequences are indicated in blue.

Supplementary Table S2. Oligonucleotides for amplification of p15A-cm linear vectors. Lowercase letters are homology arms

Gene	Primer	sequence (5'-3')
lux	lux25-1	ttcaagtaaagcattaagatcggttGGATCCTTACGCCCCGCCCTGCCACTC
	lux25-2	aaaatctaaatcaagtaataaataaGGTACCGGTAACGAATCAGACAATTGACG
lux	Lux40-1	aatttccattgaacattcaagtaaagcattaagatcggttGGATCCTTACGCCCCGCCCTGCCACTC
	Lux40-2	aataaatagatataaaaaatctaaatcaagtaataaataaGGTACCGGTAACGAATCAGACAATTGACG
lux	Lux80-1	ctaaatcgagcatttccatcactttgtaattcatcagaccaatttccattgaacattcaagtaaagcattaagatcggttGGATCCTTACGCCCCGCCCTGCCACTC
	Lux80-2	tatttttaacgagttagtagaacgctaattccgtcagaataataaatagatataaaaaatctaaatcaagtaataaataaGGTACCGGTAACGAATCAGACAATTGACG
plu2670	plu2670-1	ctgggaaaactacggtgagcatatgtttaccacttcttcagagaaccgtgaatattgcgttaaaccgatgaactgcccgggTTACGCCCCGCCCTGCCACTC
	plu2670-2	aaatccatgattaccaacccactcccaaagatattataaaatctcaacatgctgatttgatcctttggaatgggctaaatctagaGGTAACGAATCAGACAATTGACG
Mouse *Prkar1a*	mPrkar1a-1	agatgacgccctctccactctgcataccaattccacataaccactcttcaaattataatgctcacagaccctctaaggttTTACGCCCCGCCCTGCCACTC
	mPrkar1a-2	catagtgacccattgttatagagtacttagctatgcctcagagtaatgctcttgatctaggtgctttctttttacttgttGGTAACGAATCAGACAATTGACG
Mouse *Dpy30*	mDpy30-1	ggattttataagtgaagggtgacatttgtagtattccttaccccgtctggttcctcatacgcattaactcataatccttggtaccTTACGCCCCGCCCTGCCACTC
	mDpy30-2	tgccctttcagaagacctaggtttgattcctagcacccactgatgctaacaaccagacataatgccagttctaggggatccGGTAACGAATCAGACAATTGACG
Mouse *Wnt4*	mWnt4-1	gtttaactcccttaacacacacacacacacatgcacacgcacatgcacacgcgcactcacaccacaaaatcacacaatttTTACGCCCCGCCCTGCCACTC
	mWnt4-2	ctgtgctcagagctttgagtgccctacatataacaggacacaaactcttccttactttctgattgccacatggtccatttGGTAACGAATCAGACAATTGACG
Mouse *Lmbr1l-Tuba1a*	mLT-1	ctatgccataggaatttggaggaaattaggcttactggtccattaattatttccaacggatgcagtcagaatgtatatttTTACGCCCCGCCCTGCCACTC
	mLT-2	ttagagcagacagtgctcttaaccgttgagccatctctccagtccctggacacactgattttaagttgacattaaaatttGGTAACGAATCAGACAATTGACG
Human *DPY30*	hDpy30-1	ccaccacgccaggctaattttagcttaataccctattacctctcaggtgtgtctaaatttttctttgatggaaaaactagtTTACGCCCCGCCCTGCCACTC
	hDpy30-2	agttatgtcagtttgctattataatgtttattattattccttgaataaaaaagaatatgcacctaataaacgttgactagtGGTAACGAATCAGACAATTGACG
Human *IGFLR1-LIN37*	hIL-1	agtggccctagagggtcggtaaggagccaagggcctatgcaaggatgctcacacacctcccatccccaccttcccagacacatatgTTACGCCCCGCCCTGCCACTC
	hIL-2	gtgctgtccaagcttggctcatctggggtttgctgggcttaacacccaataaagaactttgctgactactaagcccagtaGGTAACGAATCAGACAATTGACG

Supplementary Table S2 continued. Oligonucleotides for amplification of p15A-cm linear vectors. Lowercase letters are homology arms

Gene	Primer	sequence (5'-3')
Human *IGFLR1-ARHGAP33*	hIA-1	acggccggcgactgctcctcagtgaggaggcgtcactcaatatccctgcagtggcggccgcccatgtgatcaaacggtaTTACGCCCCGCCCTGCCACTC
	hIA-2	gtgctgtccaagcttggctcatctggggtttgctgggcttaacacccaataaagaactttgctgactactaagcccagtaGGTAACGAATCAGACAATTGACG
Human *ZBTB32-LIN37*	hZL-1	cctagagggtcggtaaggagccaagggcctatgcaaggatgctcacacacctcccatccccaccttcccagacacatatgTTACGCCCCGCCCTGCCACTC
	hZL-2	aagaactagactagaagttggaaaacaggtttcactgcatcccttcacagtgaatggcccttctcttcagcactcatatgGGTAACGAATCAGACAATTGACG

Supplementary Table S3. Oligonucleotides used for ExoCET genotyping

Name	sequence (5'-3')
Oct4-Venus-1	ggtgacgagaacaggaacagacttgtagtcagaattaacatgatctggctggctgttcaatttttcaggggacacagggttaattaaTTACGCCCCGCCCTGCCACTC
Oct4-Venus-2	cgcccacaggtgtacagtagcaggcatgcacacaccacaaaaaaaaaaaaaaatttttttttttcaaagttgaaacagggatGGTAACGAATCAGACAATTGACG
Nanog-Cherry-1	cttggcttacatgaacagttttaaaggtgaaattaacttaagttcctattgcacaattgggtctggaacccagggccccacatatgTTACGCCCCGCCCTGCCACTC
Nanog-Cherry-2	tgggtaagagcatccaacggctcagtgggtaagagcacccaactgctcttccgaaggtctggagttcaaatcccagcaaccatatgGGTAACGAATCAGACAATTGACG
Gata2-Venus-1	ccatctccagcaggttcctaggtcctgtgcctatctggatcttggggacattgagtcttgacaataccatacatattgtaTTACGCCCCGCCCTGCCACTC
Gata2-Venus-2	cacttttccccaaagtcacctggagtttttctccacaacacctccattctggagagacaagtgagttcagtggcgagtaGGTAACGAATCAGACAATTGACG
Mll4-AID-1	cattgtagtttctttccataggccaatccttgtcagtgtaattcagtccgttgtgtaaatgcagactgtagaaaagacttactagtTTACGCCCCGCCCTGCCACTC
Mll4-AID-2	cgaaaatcttctaaacctgtcttttatttcttcttcttcttccttaaactaagctaaatactctgctacttggttggaatGGTAACGAATCAGACAATTGACG

Lowercase letters are homology arms.

Supplementary Table S4. Oligonucleotides used to amplify the *mVenus-PGK-neo* cassette

Name	**sequence (5'-3')**
mVenus-mDpy30-1	tcctagcatcttatcttttaaaaaacaaggcgcagtttgaagatcgaaatGAAGTGCATACCAATCAGGAC
mVenus-mDpy30-2	atcatgtaaatctacagtagcaaccaaatgtttcttctcttcccaagaaaCTATTTAGTCAAGCGCTCTTGTCGTCGTCATCCTTGTAG
mVenus-hDPY30-1	ttctagcatcttatcttttaaaaaacaaggcacagtttgaagatcgaaacGAAGTGCATACCAATCAGGAC
mVenus-hDPY30-2	atcatgtaaatctacagtagcaactaaatttttctgttcttcccattaagCTATTTAGTCAAGCGCTCTTGTCGTCGTCATCCTTGTAG

Lowercase letters are homology arms.

Supplementary Table S5. Sequence analysis of *Wnt4* alleles cloned from the mouse melanoma B16 cell

Position on *Mus musculus* strain C57BL/6J chromosome 4 (NC_000070.6)	137,277,158	Between 137,280,611 and 137,280,612	137,287,058	Between 137,291,949 and 137,291,950	Between 137,297,023 and 137,297,024	137,313,617
Base of NC_000070.6	C	-	G	-	-	A
#31	G	AGATAGATAGATAGAT	-	-	-	C
#71, #82	C	AGATAGATAGAT	-	AC	GAAAGAAA	A
		intron	intron	intron	exon	

*Location of *Wnt4* on chromosome 4: 137,277,489-137,299,726

Supplementary Table S6. Titration of ExoCET genotyping experiment using Oct4-hVenus #7 (R1) cell line

500 ng of digested genomic DNA mixed with indicated amounts of p15A-cm vectors		
p15A-cm vectors	**c.f.u. per ml on Km+Cm plates**	**Correct / checked***
500 ng	12	3/3, 100%
1000 ng	10	3/3, 100%
2000 ng	8	3/3, 100%
1000 ng of digested genomic DNA mixed with indicated amounts of p15A-cm vectors		
p15A-cm vectors	**c.f.u. per ml on Km+Cm plates**	**Correct / checked***
500 ng	10	3/3, 100%
1000 ng	10	3/3, 100%
2000 ng	10	3/3, 100%

* Confirmed by restriction analysis (**Supplementary Figure S7**)

Supplementary Table S7. Direct cloning of the 45kb segment containing the *Wnt4* gene from mouse genomic DNA using indicated methods

	T4pol+RecET	Gibson	Gibson+RecET
c.f.u. (/ml)	76±16	181,000±1,000	257,000±17,000
Correct/checked	8/25	0/60	0/60

Supplementary Table S8. Summary of previous direct DNA cloning methods

	Methods	Features	References
In vivo	**TAR**	Relies on the natural homologous recombination of the yeast *Saccharomyces cerevisiae.*	(1-3)
	Advantages: It can directly clone very large DNA fragments from eukaryotic and prokaryotic genomes. **Disadvantages:** 1) TAR cloning is laborious; 2) In yeast, recombination is continuously active resulting in stability problems with cloned DNAs; 3) The final recombinant DNA molecule needs to be reintroduced into *E. coli* for downstream analytical and preparative work.		
	RecET	Based on the linear-linear homologous recombination mediated by the full-length RecET in *E. coli.*	(4)
	Advantages: 1) Simple and fast; 2) The expression of RecET recombinases in *E. coli* is tightly regulated under an inducible promoter. Thereby the potential for recombination is limited to a short window, which reduces unwanted rearrangements and enhances the stable maintenance of DNA molecules with repetitive sequences; 3) The target region is cloned in well studied and tractable hosts which facilitate the further engineering; 4) Direct cloning to establish an expression construct is possible. **Disadvantages:** It is challenging to clone very large DNA fragments (>50kb) from prokaryotic genomes or to clone DNA from complex genomes.		
	oriT	Integration of two directly oriented *oriT* sites flanking the target DNA region in original hosts.	(5)
	ΦBT1	Integration of mutated *attB* and *attP* sites flanking the target DNA region in original hosts.	(6)
	Cre	Integration of two loxP sites flanking the target DNA region in original hosts.	(7)
	Advantage: Both methods can selectively isolate large fragments from prokaryotic genomes. **Disadvantages:** 1) They are limited to bacteria in culture with well-developed genome engineering systems; 3) The process is laborious employing two plasmid constructions and two genome modifications followed by a final step of DNA excision and circularization by conjugation or site-specific recombination for each exercise.		
	BGM vector	Isolated target DNA is transformed into *Bacillus subtilis* and integrated into the genome by homologous recombination using two preinstalled sequences flanking the target DNA region.	(8)
	Advantage: The capacity of *Bacillus subtilis* genome vector is large (>3.5 Mb). **Disadvantages:** 1) Pre-engineering of *Bacillus subtilis* genome is required; 2) The utilization of the cloned DNA in the BGM vector is inconvenient.		

	Methods	Features	References
In vitro	**Gibson** **CATCH**	The linear cloning vector and digested genomic fragment are specifically annealed using the *in vitro* Gibson assembly reaction.	(9,10)
	Advantage: Gibson assembly is an easy *in vitro* reaction. **Disadvantage:** High background from re-circularized empty vector (see Supplementary Table S7).		

Supplementary References

1. Larionov, V., Kouprina, N., Graves, J., Chen, X.N., Korenberg, J.R. and Resnick, M.A. (1996) Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 491-496.
2. Kouprina, N. and Larionov, V. (2008) Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protoc.*, **3**, 371-377.
3. Lee, N.C., Larionov, V. and Kouprina, N. (2015) Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast. *Nucleic Acids Res.*, **43**, e55.
4. Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P.M., Plaza, A., Xia, L., Muller, R., Stewart, A.F. *et al.* (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.*, **30**, 440-446.
5. Chain, P.S., Hernandez-Lucas, I., Golding, B. and Finan, T.M. (2000) oriT-directed cloning of defined large regions from bacterial genomes: identification of the *Sinorhizobium meliloti* pExo megaplasmid replicator region. *J. Bacteriol.*, **182**, 5486-5494.
6. Du, D., Wang, L., Tian, Y., Liu, H., Tan, H. and Niu, G. (2015) Genome engineering and direct cloning of antibiotic gene clusters via phage varphiBT1 integrase-mediated site-specific recombination in Streptomyces. *Sci. Rep.*, **5**, 8740.
7. Hu, S., Liu, Z., Zhang, X., Zhang, G., Xie, Y., Ding, X., Mo, X., Stewart, A.F., Fu, J., Zhang, Y. *et al.* (2016) "Cre/loxP plus BAC": a strategy for direct cloning of large DNA fragment and its applications in *Photorhabdus luminescens* and *Agrobacterium tumefaciens*. *Sci. Rep.*, **6**, 29087.
8. Yonemura, I., Nakada, K., Sato, A., Hayashi, J.I., Fujita, K., Kaneko, S. and Itaya, M. (2007) Direct cloning of full-length mouse mitochondrial DNA using a *Bacillus subtilis* genome vector. *Gene*, **391**, 171-177.
9. Jiang, W., Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y. and Zhu, T.F. (2015) Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun.*, **6**, 8101.
10. Zhou, Y., Murphy, A.C., Samborskyy, M., Prediger, P., Dias, L.C. and Leadlay, P.F. (2015) Iterative mechanism of macrodiolide formation in the anticancer compound conglobatin. *Chem. Biol.*, **22**, 745-754.