Polaris: A toolkit for standardized construction, integration, and quantitative testing in *Escherichia coli* and *Saccharomyces cerevisiae*

Ronghui Xie [a,1], Ran Wang [a,1], Yi Zhan [a,c], Boyan Wei [a], Siyang Xu [a], Ye Chen [a,b,*], Lili Liu [a,d,*]

[a] Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Avenue, Shenzhen University Town, Nanshan, Shenzhen 518055, China.
[b] University of Chinese Academy of Sciences, 1 Yanqihu East Rd, Beijing, 101408, China
[c] College of Biological Sciences, China Agricultural University, No. 2 Yuanmingyuan West Road, Haidian District, Beijing 100193, China.
[d] Faculty of Synthetic Biology, Shenzhen University of Advanced Technology, No. 1 Gongchang Road, Guangming District, Shenzhen, Guangdong 518107, China.
[1] These authors contributed equally to this work.
[*] Corresponding authors.
E-mail addresses: liulili@suat-sz.edu.cn (L. Liu); chenye.h.ai@gmail.com (Y. Chen)

Abstract

Existing synthetic biology toolkits have expanded part libraries and assembly capabilities. However, new users may still encounter ambiguity in construct definitions, hierarchical transitions, genomic integration routes, and quantitative testing procedures. Here, we present Polaris, a Golden Gate toolkit for *Escherichia coli* and *Saccharomyces cerevisiae*. Polaris defines a shared three-level hierarchy. Level 1 stores reusable parts, Level 2 generates transcription units and other intermediate constructs, and Level 3 supports multigene assembly and switching between testing and integration backbones. The toolkit distinguishes part interfaces from transcription unit interfaces, uses BpiI and BsaI in successive assembly stages, and provides parallel ccdB and mScarletI3 screening options. Integration modules and quantitative characterization frameworks were implemented in both hosts and demonstrated using representative genetic parts and activation and repression systems. A companion digital platform further encodes the Level 1 to Level 3 rules for assembly planning. Polaris connects hierarchical construction, stable genetic implementation, and quantitative testing within a common execution framework, providing a practical basis for data accumulation, part standardization, and future automation.

Keywords: Synthetic biology toolkit; Golden Gate assembly; Genetic part standardization; Hierarchical DNA assembly; Genomic integration; Quantitative characterization

1. Introduction

The engineering implementation of synthetic biology relies on the standardization, modularization, and reusable assembly of genetic parts. Golden Gate assembly, together with the Modular Cloning (MoClo) framework based on Type IIS restriction enzymes, has become a central foundation for standardized part storage, hierarchical assembly, and automation-compatible execution [1–7]. These approaches enable basic genetic parts to be maintained in predefined formats and subsequently assembled in one-pot reactions through standardized interfaces. Their utility has been demonstrated across diverse organisms through toolkits such as the Yeast Toolkit, CIDAR MoClo, EcoFlex, Golden Standard, and the Coli Toolkit [3–7]. As high-throughput construction and design–build–test workflows continue to develop, hierarchical assembly schemes have also become an important basis for part reuse, automated construction, data accumulation, and database-oriented management [2,8–11].

Escherichia coli and *Saccharomyces cerevisiae* are among the most mature and experimentally tractable prokaryotic and eukaryotic chassis, respectively [3,5,7]. Both organisms benefit from well-established genetic manipulation methods, extensive collections of characterized parts, and broadly accessible cultivation and measurement procedures. Together, they provide representative platforms for the two major cellular expression architectures used in synthetic biology. A construction framework covering both hosts can therefore support methodological transfer between prokaryotic and eukaryotic engineering workflows, while allowing host-specific differences in transcription, translation, plasmid maintenance, and genomic integration to be retained. This is particularly useful for laboratories that need to establish working pipelines rapidly and subsequently expand them toward different hosts or application contexts.

As DNA assembly technologies have matured, an important engineering challenge has shifted from whether fragments can be assembled successfully to whether construct objects, hierarchical transitions, and downstream usage routes are defined with sufficient clarity. For new users, the main learning effort often lies not in performing the restriction–ligation reaction itself, but in determining which level a part belongs to, how

interfaces are partitioned, which enzymes and backbones are required at each stage, and how completed constructs are transferred into stable genetic backgrounds and testing frameworks. Existing toolkits provide extensive resources for part expansion, host adaptation, and complex multigene construction [1–7]. However, laboratories adopting such systems may still need to coordinate several separate decisions before beginning routine experiments, including part domestication, hierarchical upgrading, backbone selection, clone screening, genomic integration, and the selection of appropriate quantitative testing architectures.

Quantitative characterization further reinforces the need to connect these steps. A comparable genetic object is not defined by sequence alone, but by its sequence together with its assembly interfaces, vector context, host background, reference construct, cultivation condition, and measurement procedure. Reference frameworks such as relative promoter units and relative expression units illustrate how fixed reference constructs and readout logic can make quantitative measurements more interpretable and comparable [12–14]. At the same time, genetic context, local sequence junctions, host physiology, and environmental conditions can influence the measured performance of synthetic biological systems [15,16]. A toolkit intended for routine engineering should therefore provide more than an assembly syntax. It should also define clear routes through which assembled constructs can enter stable genomic contexts and standardized testing configurations, allowing construction objects to be converted into measurable performance objects [11,17,18].

To address this need, we developed Polaris, a dual-host Golden Gate toolkit for *E. coli* and *S. cerevisiae*. Polaris uses a shared three-level object hierarchy that connects reusable part storage, transcription-unit assembly, multi-transcription-unit construction, genomic integration, and quantitative testing. Level 1 contains reusable basic parts; Level 2 generates transcription units and other intermediate constructs that can be further upgraded; and Level 3 supports multi-transcription-unit construction and switching among plasmid-testing, genomic-integration, and host-specific application backbones. The two host implementations follow the same hierarchical progression and transcription-unit assembly logic while retaining host-specific part architectures, backbones, integration

mechanisms, and testing procedures.

The toolkit distinguishes part-level interfaces from transcription-unit-level interfaces and uses BpiI and BsaI hierarchically to make the roles of different assembly stages explicit [2,5,6]. It also provides parallel ccdB- and mScarletI3-based backbone versions, enabling either growth-based counterselection or direct visual screening according to laboratory requirements [19–21]. In addition, Polaris incorporates host-specific genomic-integration modules, defined architectures for part and regulatory-system characterization, and a companion digital platform for plasmid-map visualization and hierarchical assembly management. The principal contribution of Polaris therefore does not lie in introducing a new Type IIS assembly chemistry or merely expanding a part collection, but in connecting a shared three-level object hierarchy with host-specific integration routes, quantitative testing architectures, dual screening options, and digitally encoded assembly workflows in two widely used microbial chassis. Here, we describe the design and implementation of the toolkit and demonstrate its application to hierarchical construction, genomic integration, genetic-part characterization, regulatory-system testing, and multilevel digital assembly planning.

2. Materials and methods

2.1. Strains and plasmids

The *E. coli* cloning strains used in this study were DH10B, DH10B-Alt (λ attB::*lacI*, *tetR*, *araC*), DH10B-12S [22], and DB3.1. The *E. coli* strains used for system characterization were MG1655-Alt (MG1655 λ attB::*lacI*, *tetR*, *araC*) and ECint034 [23]. The *S. cerevisiae* strains used for characterization were BY4741 and CYE72. Detailed information on the test plasmids and strains is provided in Supplementary Tables 2 and 3.

2.2. Media and reagents

LB liquid medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl); LB agar medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 20 g/L agar); 2×YT medium (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl); TSS solution (85% liquid LB, 10% PEG [wt/vol, MW8000], 5% DMSO [vol/vol], 50 mM $MgCl_2$ [pH 6.5]; LB liquid medium and $MgCl_2$ were sterilized by autoclaving, whereas PEG and DMSO were sterilized by filtration); 5×KCM

(0.5 M KCl, 0.15 M $CaCl_2$, 0.25 M $MgCl_2$); M9 medium supplemented with 0.4 % glycerol, 1 % 100× trace element and 0.2 % Casamino acids; 100× trace element solution (1.0 g/L $FeSO_4$•$7H_2O$, 1.2 g/L $MnCl_2$•$4H_2O$, 0.8 g/L $CoCl_2$•$6H_2O$, 0.7 g/L $ZnSO_4$•$7H_2O$, 0.3 g/L $CuCl_2$•$2H_2O$, 20 mg/L H_3BO_3, 0.25 g/L $(NH_4)_6Mo_7O_{24}$•$4H_2O$, 5.0 g/L EDTA); SD/-Leu (Coolaber, PM2201); SD/-Ura (Coolaber, PM2271); SD/-Ura-Leu (Coolaber, PM22991); SMSG-Ura (Coolaber, T2001); SD-All medium (Coolaber, PM2101); and agar (Bacto, 214010). The working concentrations of antibiotics were as follows: kanamycin (50 μg/mL, Aladdin, K103024), chloramphenicol (25 μg/mL, Sangon Biotech, A100230), gentamicin (10 μg/mL, BBI, A620217), carbenicillin (100 μg/mL, Macklin, C805408-10g), and G418 (200 μg/mL, Coolaber, CG5471). Information on the inducers used in the characterization experiments is provided in Supplementary Table 4.

2.3. Plasmid construction by Gibson assembly

Backbone plasmids were constructed using Gibson assembly (CU201, TransGen Biotech). Linearized vectors were obtained by restriction enzyme digestion or PCR amplification. Each assembly fragment shared a 15–25 nt overlap with the linearized vector. The reaction mixture consisted of 5 μL 2× Basic Assembly Mix, 5–100 ng linearized vector (recommended: approximately 0.03 pmol), each insert fragment at 0.03–0.06 pmol (vector: insert molar ratio of 1:1 to 1:2), and ddH_2O added to a final volume of 10 μL. After gentle mixing, the reaction was incubated at 50 °C for 5–15 min. The reaction products were used for transformation or stored at −20 °C.

2.4. Plasmid construction by Golden Gate assembly

A standard Golden Gate reaction mixture contained DNA fragments at appropriate volumes (the amount of each fragment was adjusted proportionally, with an insert-to-backbone molar ratio of approximately 2:1, and approximately 0.05 pmol backbone plasmid), 0.5 μL BpiI (Thermo Scientific, ER1012) or BsaI-HF®v2 (NEB, R3733L), 0.5 μL T4 DNA ligase (NEB, M0202L), 1.0 μL 10× T4 DNA ligase buffer, and nuclease-free water (ddH_2O) to a final volume of 10 μL. The reaction program was as follows: (37 °C for 2 min, 16 °C for 2 min) for 25–35 cycles. When the number of fragments exceeded six or when complex assemblies were performed, the cycling conditions were optimized to (37 °C for 5 min, 16 °C for 5 min) for 35 cycles. After cycling,

the reaction was incubated at 50 °C for 15 min to inactivate T4 DNA ligase while maintaining restriction enzyme activity to digest the template. If transformation was not performed immediately, the reaction was further incubated at 80 °C for 20 min to fully inactivate the restriction enzyme.

2.5. Preparation of competent *E. coli* cells

Strains were streaked on LB agar plates and incubated overnight at 37 °C. A single colony was inoculated into 2 mL LB liquid medium and grown overnight at 37 °C, 220 rpm. Then, 1 mL of the culture was transferred into 100 mL LB medium in a 500 mL flask and grown at 37 °C, 220 rpm for 2–3 h until OD_{600} = 0.3–0.35. The flask was chilled on ice for 20 min, and the cells were harvested by centrifugation at 4 °C, 1000 × g for 15 min. The supernatant was discarded, and the pellet was resuspended in 4 mL pre-chilled TSS solution. This washing step was repeated twice. Finally, 100 μL aliquots of competent cells were dispensed into 0.2 mL tubes and stored at −80 °C.

2.6. Chemical transformation of *E. coli*

Competent cells were thawed on ice, and an appropriate amount of competent cells (10–50 μL) was transferred into a 1.5 mL microcentrifuge tube (for TSS competent cells, 5×KCM was added to a final concentration of 1×). Plasmid DNA or assembly products were added (not exceeding 10 % of the competent cell volume), mixed gently by flicking 4–5 times, and incubated on ice for 30 min. The cells were heat-shocked at 42 °C for 30 s, immediately chilled on ice for 2 min, and then supplemented with 5–10 volumes of 2×YT medium. After recovery at 37 °C, 220 rpm for 1 h, the cells were centrifuged at 600–800 × g for 5 min, part of the supernatant was removed, and the pellet was resuspended in approximately 100 μL of the remaining medium. The cells were spread onto LB agar plates containing the appropriate antibiotics and incubated overnight at 37 °C. Backbone plasmids were transformed into the DB3.1 cloning strain (carrying the gyrA462 gene), whereas non-backbone plasmids were transformed into DH10B or DH10B-Alt strains.

2.7. Genomic integration in *E. coli*

After Golden Gate assembly, backbone plasmids containing attP were digested with BsaI (or AvrII) and SpeI to remove the pSC101 origin of replication. The target fragment was isolated by 0.8% agarose gel electrophoresis and purified. The recovered fragment

was ligated with T4 DNA ligase at 25 °C for 2 h (or overnight at 16 °C) and then heat-inactivated at 65 °C for 15 min. Subsequently, BsaI (or AvrII) and SpeI were added for digestion at 37 °C for 30 min, followed by inactivation at 80 °C, yielding the circularized product.

Competent cells of the chassis strain were prepared using the TSS method and transformed with the temperature-sensitive plasmid pEcint01, and transformants were obtained at 30 °C. A single colony was inoculated into LB liquid medium containing carbenicillin and grown overnight at 30 °C with shaking. The culture was diluted 1:100 into fresh LB containing carbenicillin and grown at 30 °C until OD_{600} ≈ 0.3–0.35, followed by rapid transfer to 37 °C for 15 min to induce integrase expression. Competent cells were prepared again using the TSS method and transformed with the circularized product, and transformants were obtained at 30 °C. Colony PCR was used to verify genomic integration in the transformants. A single verified colony was picked and inoculated into antibiotic-free LB medium, followed by overnight incubation at 37 °C with shaking to cure the temperature-sensitive pEcint01 plasmid.

Competent cells were then prepared from correctly integrated strains and transformed with the pCP20 plasmid, followed by incubation at 30 °C. Single colonies were streaked onto carbenicillin plates and incubated at 30 °C for 3–6 h, and PCR was used to confirm the removal of the kanamycin resistance gene. KanR-negative clones were inoculated into antibiotic-free LB and grown at 37 °C for 2–3 h, then streaked onto plates with and without carbenicillin and incubated overnight at 37 °C. Clones that failed to grow on carbenicillin plates were considered to have successfully lost the pCP20 plasmid.

2.8. Characterization of *E. coli* parts and transcriptional regulatory system

In 96-well plates (Costar 3799), 200 μL LB medium (with or without the appropriate antibiotics or inducer) was added to each well. Single colonies were inoculated into the wells and cultured overnight at 37 °C, 800 rpm in a thermostatic shaker (Biosan, PST-60HL-4). Subsequently, in another 96-well plate, 100 μL M9 medium was added to each well (with or without antibiotics; if transcription factor expression required induction, the corresponding inducer was included). Then, 1 μL of overnight culture was transferred into each well and incubated at 37 °C, 800 rpm for 2.0 h (first-stage culture). Afterward,

100 μL M9 medium was added to each well (with or without antibiotics; transcription factor induction conditions were the same as in the first stage; if the transcription factor required its cognate ligand, the ligand was added at this stage), and the culture was continued at 37 °C, 800 rpm for 3.0 h (second-stage culture). Finally, the OD_{600} and corresponding fluorescence values of each well were measured using a microplate reader.

2.9. Transformation of *S. cerevisiae*

In the classical HR integration method, 5 μg of plasmid was digested with BsaI, and the reaction product was directly transformed into prepared competent cells. For the CRISPR/Cas9 integration method, in addition to the above enzyme-digested product, 200 ng each of the Cas9 plasmid (pCfB2312) and the sgRNA plasmid (sgRNA+pScCom02) were also required. After verification of correct integration in yeast by PCR or sequencing, the sgRNA plasmid needed to be eliminated through a curing procedure. Overnight cultures of the verified yeast strains were diluted 10,000-fold and spread onto YPD plates supplemented with 5-fluoroorotic acid (HARVEYBIO, 5-FoA) at 1 mg/mL. The plates were incubated inverted at 30 °C for 2 days to obtain transformants. Individual colonies that grew out were picked and tested for plasmid loss by streaking on YPD plates containing G418 (200 μg/mL), inoculating into liquid SD/-Ura medium, and streaking on SD/-Ura plates. Failure to grow after 2 days under all three conditions indicated that both the Cas9 and sgRNA plasmids had been successfully lost. The resulting strains were then preserved and stored for subsequent experiments.

Yeast competent cells were prepared and transformed using the Classic Yeast Competent Cell Preparation and Transformation Kit (Frozen type) (Coolaber, SK2400). Transformation products were spread onto the corresponding selective SD plates and incubated upside down at 30 °C for 2 days to obtain transformants.

Alternatively, the Frozen EZ Yeast Transformation Kit II (Zymo Research Corporation, T2001) was used for yeast competent cell preparation and transformation. Transformation products were spread onto the corresponding selective SD plates and incubated upside down at 30 °C for 2 days to obtain transformants.

2.10. Characterization of *S. cerevisiae* parts and transcriptional regulatory system

For characterization of constitutive and inducible promoter elements, single colonies

were inoculated into 96-deep-well plates (Yderbio, YDASQ01-S) containing 500 μL selective SD medium per well and cultured at 30 °C, 800 rpm for 24 h. The cultures were then diluted 1:200 into fresh SD medium supplemented with the appropriate inducer and cultured at 30 °C, 800 rpm for another 16 h. Then, 20 μL of culture was mixed with 180 μL phosphate-buffered saline (PBS, Proteintech, PR20014) in a 96-well U-bottom plate to prepare samples for flow cytometry. For terminator characterization, transformants were picked into the corresponding selective medium containing 10 mmol/L D-xylose. After 24 h of cultivation, the cultures were diluted 1:200 into SD medium containing 10 mmol/L D-xylose and grown at 30 °C, 800 rpm for 16 h, after which samples were collected for flow cytometry. Sample preparation was the same as that used for promoter characterization.

For transcription factor quantification, single colonies were inoculated into 96-deep-well plates containing 500 μL selective SD medium per well and cultured at 30 °C, 800 rpm for 24 h. The culture was then diluted 1:200 into fresh SD medium containing a gradient of TF expression inducer concentrations, with parallel non-induced and induced control groups supplemented with either no or saturating concentrations of the TF effector, respectively. After further incubation for 16 h, cells were harvested and subjected to flow cytometric analysis as described above. The Input RPU values obtained from the input module and the Output RPU values from the reporter module were used to generate Input RPU–Output RPU curves, from which the TF input condition yielding the maximal output difference between the two groups was identified. Subsequently, with this TF input intensity fixed, a gradient of TF effector concentrations was applied, and the same cultivation and detection procedures were followed.

2.11. Microplate reader measurements (for *E. coli*)

A microplate reader (Tecan Infinite 200 PRO E Plex) was used to measure the OD_{600} and corresponding fluorescence channels for each sample well. Appropriate fluorescence gain settings were selected to obtain suitable signal values. The microplate reader was configured as follows: mTagBFP2, Ex/Em: 420/455 nm; sfCFP, Ex/Em: 433/470 nm; sfGFP, Ex/Em: 480/510 nm; sfYFP, Ex/Em: 500/530 nm; mCherry, Ex/Em: 580/610 nm. The initial data processing for each sample was performed according to the following

formula:

$$\text{Flu/OD（sample）} = {(\text{Flu}_{(\text{sample})} - \text{Flu}_{(\text{control})})}/{(\text{OD}_{(\text{sample})} - \text{OD}_{(\text{medium})})} \quad \text{A.1}$$

Here, $Flu_{(sample)}$ and $Flu_{(control)}$ are fluorescence readings of the test sample and blank-cell control, respectively, while $OD_{(sample)}$ and $OD_{(medium)}$ are the corresponding OD_{600} readings of the test sample and culture medium, all measured by the microplate reader.

2.12. Fluorescent protein spectral scanning

Fluorescent protein spectra were characterized using an Infinite 200 PRO E Plex multimode microplate reader (Tecan). The OD_{600} of each culture was measured immediately before fluorescence acquisition to monitor cell density, and spectral measurements were subsequently performed in the same wells. For the initial characterization in *E. coli*, paired excitation/emission wavelength scans were performed for mTagBFP2, sfCFP, sfGFP, sfYFP, mCherry, and mScarletI3. Excitation wavelengths were varied over protein-appropriate ranges between 400 and 650 nm using increments of 5 or 10 nm. At each excitation wavelength, the emission wavelength was set 30 nm above the excitation wavelength (Em = Ex + 30 nm). The resulting signal-to-background ratios (SBRs) were used to generate the paired excitation/emission scan curves. Additional sequential scans were performed for sfCFP and mTagBFP2 to refine their excitation and emission settings. For sfCFP, the excitation wavelength was first fixed at 430 nm while the emission wavelength was scanned from 460 to 550 nm using increments of 5 or 10 nm. The emission wavelength was then fixed at 470 nm while the excitation wavelength was scanned from 400 to 440 nm using increments of 3, 5, or 10 nm. A second excitation scan was performed from 400 to 445 nm using increments of 5 or 10 nm with the emission wavelength fixed at 475 nm. For mTagBFP2, the excitation wavelength was fixed at 420 nm while the emission wavelength was scanned from 450 to 520 nm using increments of 5 or 10 nm. The emission wavelength was subsequently fixed at 455 nm while the excitation wavelength was scanned from 400 to 425 nm in 5-nm increments. For spectral characterization in *S. cerevisiae*, paired excitation/emission

wavelength scans were performed for mTagBFP2, mCitrine, mNeonGreen, mScarletI3, and mRuby2 over protein-specific wavelength ranges spanning 255-700 nm, using increments of 5 or 10 nm. A fixed 50-nm interval between excitation and emission wavelengths was used for mTagBFP2 and mRuby2, whereas a fixed 30-nm interval was used for mCitrine, mNeonGreen, and mScarletI3. For every wavelength setting, a corresponding non-fluorescent host control was measured using the same wavelength and gain settings. Gain was adjusted as required to keep the fluorescence readings within the measurable range, but identical settings were used for the fluorescent sample and its corresponding background control at each wavelength combination. The sample and background-control signals were used to calculate the SBR at each wavelength combination.

2.13. Flow cytometry operation and analysis

Fluorescence measurements were performed using a FACSCelesta flow cytometer equipped with a high-throughput sampler (HTS, BD Biosciences). Instrument settings were optimized separately for *S. cerevisiae* and *E. coli*. For *S. cerevisiae*, the FSC, SSC, and fluorescence detector voltages were adjusted while monitoring the FSC-A, SSC-A, and fluorescence signals to position the main cell population within an appropriate region of the detection range. For *E. coli*, the SSC and fluorescence detector voltages were adjusted while monitoring the SSC-H and fluorescence signals to position the bacterial population within an appropriate detection range. An SSC threshold of 300 was applied during acquisition to suppress low-level background noise for *S. cerevisiae* and *E. coli*. More than 10,000 single-cell events were collected for each sample. Fluorescence signals were recorded in the BV421-A, BV510-A, FITC-A, and PE-Texas Red-A channels for *S. cerevisiae* and in the BV421-H, BV510-H, FITC-H, and PE-Texas Red-H channels for *E. coli*. Data were analyzed using FlowJo software (version 10; BD Biosciences), and the median fluorescence intensity of each gated cell population was extracted. Background-corrected fluorescence values were reported in arbitrary units (a.u.) and calculated by subtracting the corresponding background autofluorescence from the measured median fluorescence intensity.

2.14. Software implementation and source-code availability

The companion Polaris Kit Assembly platform was developed to support genetic-part and backbone selection, compatibility checking, hierarchical construct generation, plasmid-map visualization, and assembly-route management according to the Level 1–3 rules defined in this toolkit. The deployed web platform is available at https://infrasynbio.siat.ac.cn/kitassembly. The source code and associated resources for Polaris Kit Assembly are available at https://github.com/ChenYe-Lab/PolarisKitAssembly.

3. Results

3.1. A hierarchical toolkit design for standardized construction

The present toolkit adopts the hierarchical Golden Gate assembly framework [1-7], but its design emphasis lies not in adding more hierarchical levels or expanding functional branches. Instead, it is built to predefine construction objects, hierarchical boundaries, and screening logic in both hosts, so that users can proceed directly to part characterization, genomic integration, and regulatory-system construction without first developing a customized toolkit. To this end, we implemented a unified three-level architecture: Level 1 stores reusable basic parts; Level 2 serves as an intermediate construction state used to generate single transcription units (TUs) or other modules that can be further upgraded; and Level 3 corresponds to final application backbones, which can accommodate multiple TUs and can be efficiently switched among plasmid testing vectors, genomic integration vectors, and backbones with different host or replicon contexts through predefined interfaces. Because part characterization typically does not require full multi-TU upgrading, we additionally provide a direct Level 1-to-Level 3 route for standardized testing in order to further reduce the construction burden.

In interface design, the toolkit retains a compatible scar framework while further strengthening the definition of part boundaries (Table 1). Part-level scars and transcription-unit scars are clearly distinguished: the former are used for ordered assembly of basic parts such as promoters, ribosome binding sites, coding sequences, and terminators, whereas the latter are used for combining multiple TUs at higher levels. Key junctions serve not only as assembly elements but also as definitions of expression architecture. For example, A2 is located downstream of the promoter and close to the

transcription start region, whereas the B scar AATG directly embeds the start codon [2,4,6]. At the CDS–terminator junction, each CDS part is designed with an additional in-frame terminal TAA immediately followed by the C scar TAAA. The first three nucleotides of the C scar form a second in-frame TAA, thereby generating tandem TAA stop codons and providing redundant translation-termination signals (Supplementary Note 1). In this way, scars are not merely short ligation sequences, but explicit definitions of part and module boundaries. To reduce the influence of upstream modules on downstream testing configurations, some Level 2 backbones also include neutral spacer sequences or terminator/insulator elements between TU interfaces and part interfaces, thereby mitigating genetic context effects [15,16] (Fig. 1).

Table 1. Scar sequences for universal Golden Gate assembly in *E. coli* and *S. cerevisiae.*

Scar	Sequence(5'→3')	Scar	Sequence(5'→3')	Scar	Sequence(5'→3')
A	GTGC	I	AGGT	R	AAAG
A2	ATCA	J	CACC	S	ACAA
B	AATG	K	TGTC	T	ACTC
C	TAAA	L	CGCT	U	AGAC
D	CCTC	M	CCGT	V	CAGG
E	GCTT	N	GCGA	W	CCAC
F	CTGA	O	GGGC	X	CCCA
G	TACG	P	AGTC	Y	CGAG
H	TTCC	Q	ATTG	Z	GTTA

The scars are divided into two types: part-level scars, including A, A2, B, C, and D, and transcription-unit-level scars, including E–Z. An additional in-frame TAA is included at the end of each CDS part. Together with the TAA triplet embedded in the C scar TAAA, the assembled CDS–terminator junction contains tandem TAA stop codons.

The system uses BpiI and BsaI in a hierarchical fashion: BsaI is used for the most common Level 2 assemblies, whereas BpiI is used for Level 1 entry and Level 3 expansion. Some previously reported hierarchical toolkits also alternate Type IIS enzymes between levels [4-6], so enzyme swapping across levels is not, in itself, the defining distinction of this system. The key design choice here is the use of BpiI rather than

alternatives that require higher reaction temperatures. Because T4 DNA ligase loses activity at elevated temperatures, while Level 2 assembly is the most frequently used step and Level 3 multi-TU assembly is particularly sensitive to overall cutting and ligation efficiency, we retained the most robust and convenient enzyme, BsaI, for Level 2 and selected BpiI, whose optimal reaction temperature is 37 °C, for Level 1 and Level 3. This makes it easier to maintain the entire hierarchical workflow within a temperature window compatible with T4 DNA ligase. Domesticated parts are first introduced into Level 1 backbones via BpiI; after correct insertion, BpiI sites are removed while BsaI sites are retained. BsaI is then used to assemble Level 2 transcription units, while BpiI sites are preserved for Level 3 expansion (Fig. 1). This division of labor makes hierarchical upgrading paths and enzymatic roles explicit and is also well suited to subsequent automation and part-management workflows.

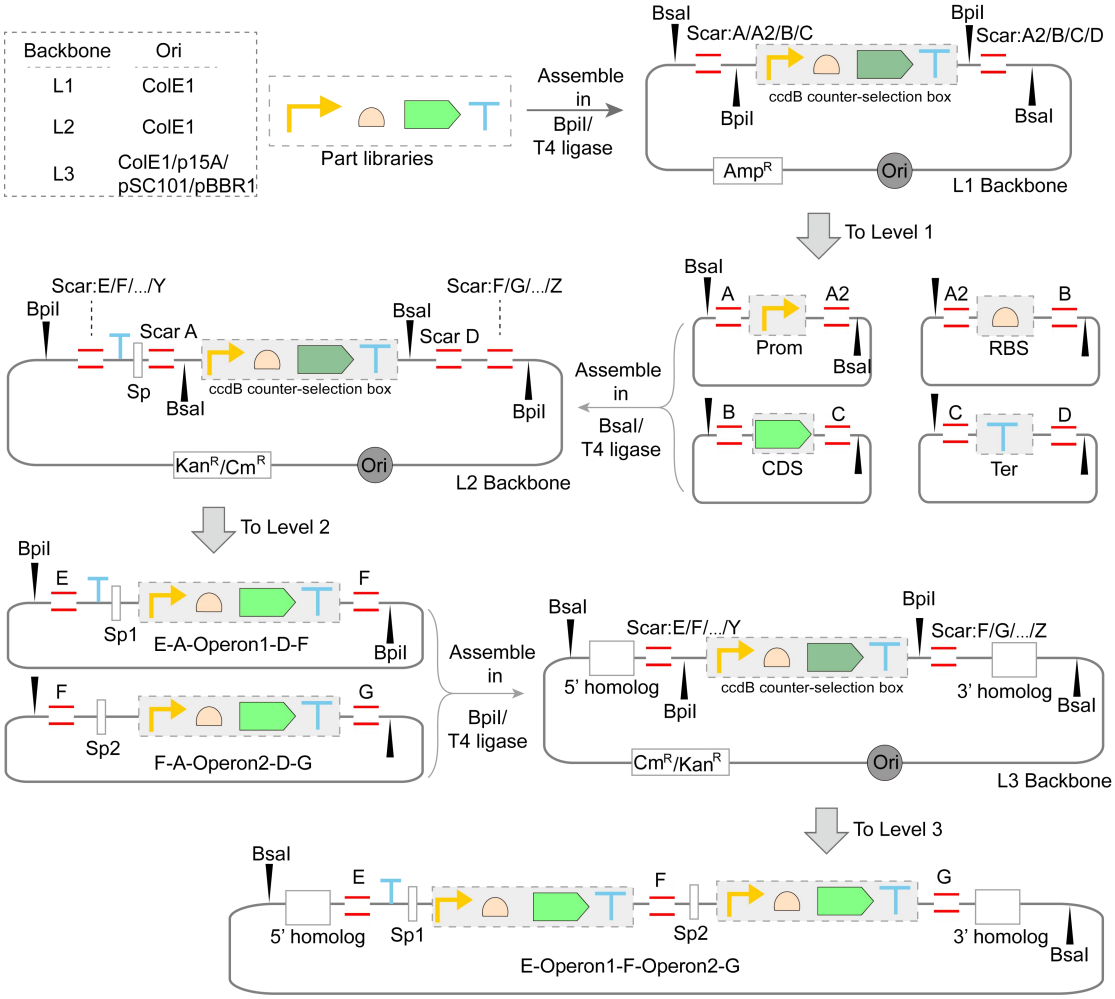


Fig. 1. Hierarchical Golden Gate assembly. In the standard Golden Gate workflow, domesticated part libraries are first used as templates to generate linear DNA fragments for Golden Gate reactions by PCR.

These fragments are then assembled into the corresponding L1 backbones through a BpiI/T4 ligase reaction to generate different types of Level 1 part plasmids. Level 1 part plasmids are further assembled into Level 2 plasmids (single transcription units) through a BsaI/T4 ligase reaction, and multiple Level 2 plasmids can subsequently be assembled into Level 3 plasmids (multiple transcription units) through a BpiI/T4 ligase reaction. The L3 backbone illustrated here represents an integration-compatible example and therefore contains flanking homology arms. These homology arms are optional, backbone-specific features rather than intrinsic components of L3 assembly. The same L3 assembly logic applies to non-integrating L3 plasmids, including many *E. coli* L3 backbones, by using backbones without homology arms. Backbones at different levels may carry different selection markers and origins of replication (Supplementary Notes 2 and 3). For clarity, only backbone plasmids using ccdB-based counterselection are illustrated. The corresponding red-fluorescent screening versions, in which ccdB is replaced by mScarletI3 (Supplementary Fig. 1), follow the same hierarchical architecture and assembly logic. Plasmids at each level have their corresponding backbone vectors, and backbones of different levels carry different selection markers or origins of replication (Supplementary Note 2, 3). Prom: Promoter, Ter: Terminator.

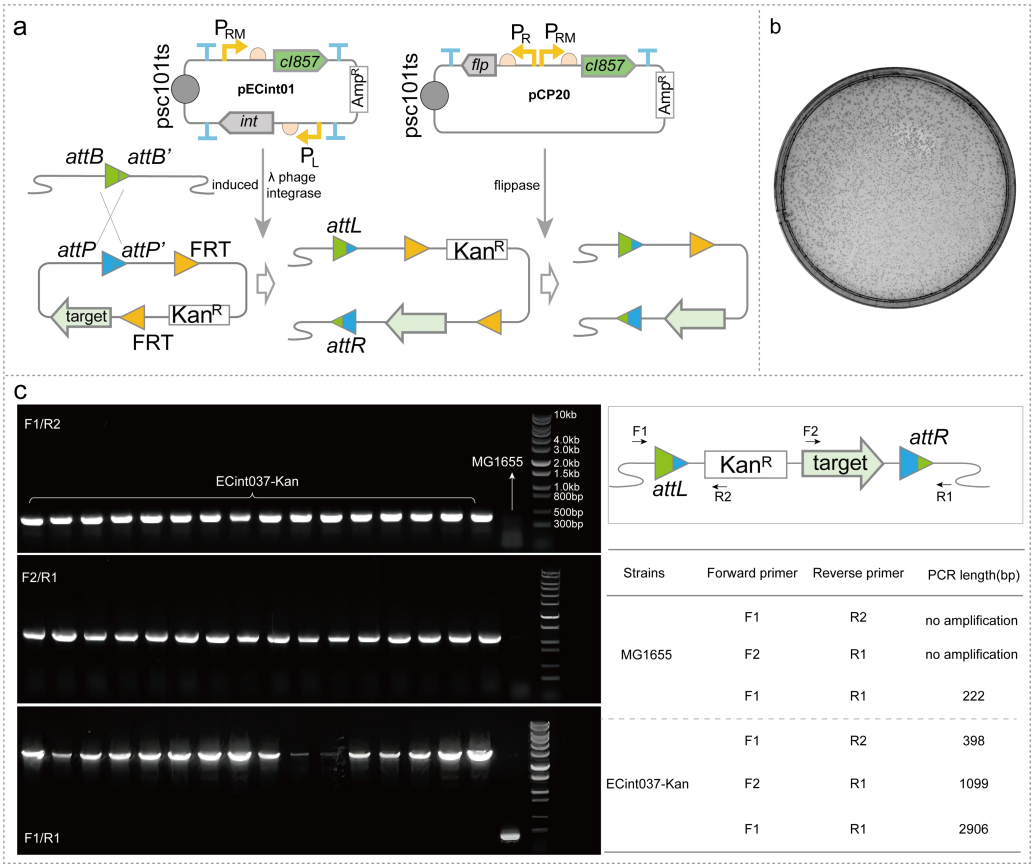
With respect to screening, all backbones are available in a ccdB version, and a parallel red-fluorescent version was generated by fully replacing ccdB with mScarletI3 (Supplementary Fig. 1). The former directly suppresses background arising from undigested backbones, self-ligation, and incorrect ligation through growth-based selection [19,20], whereas the latter enables intuitive screening in laboratories that do not maintain DB3.1 or other ccdB-compatible propagation strains [21]. These two backbone types follow exactly the same hierarchical organization and interface rules, differing only in the screening mode used under different laboratory conditions. We did not adopt the lacZ blue-white system in order to avoid the inconvenience of X-gal and the false positives associated with color-based interpretation [24,25]. Collectively, these design choices establish a dual-host construction route with clearly defined objects, explicit hierarchical progression, direct screening readout, and good compatibility across different laboratory settings.

3.2. Dual-host genomic integration modules for stable testing backgrounds

To improve the stability and comparability of downstream characterization, genomic integration is treated as a key component of the workflow in both hosts. This is especially important in *S. cerevisiae*, where single-copy genomic integration reduces the impact of plasmid copy-number variation and selection-related background differences on

quantitative measurements, thereby providing a more stable testing background [3]. In *E. coli*, genomic integration likewise provides a more stable expression background, although higher-throughput multi-site strategies typically require chassis strains preloaded with predefined integration sites [26,27]. On this basis, Level 3 backbones are designed not only to support multi-TU assembly but also to function as a unified entry point for moving constructs into stable genomic contexts.

In *E. coli*, we established a site-specific integration module based on λ phage integrase. Level 3 constructs carrying attP are processed and introduced into the genomic attB site to achieve stable targeted integration [28]. Experimentally, this system yielded more than 10^4 integration colonies in a single transformation, and all 16 tested colonies yielded band patterns consistent with the expected integration junctions (Fig. 2), confirming successful λ integrase-mediated integration among the colonies examined in this experiment. In addition to the λ system, the toolkit includes three orthogonal integrase-site systems (designated Int2, Int5, and Int7 in the source literature) [26,27], together with the corresponding landing-pad tool plasmids [29], for future multi-site genomic integration. These additional systems were not experimentally validated in this study. Such systems require chassis strains in which the corresponding integration sites have been preinstalled before the target constructs are introduced.



Strains	Forward primer	Reverse primer	PCR length(bp)
MG1655	F1	R2	no amplification
	F2	R1	no amplification
ECint037-Kan	F1	R1	222
	F1	R2	398
	F2	R1	1099
	F1	R1	2906

Fig. 2. λ phage integrase-mediated genomic integration in *E. coli*. (a) Schematic overview of the λ phage site-specific integration and marker-removal workflow. Donor circular DNA carrying an attP site and the target cassette is recombined with the chromosomal attB site under the action of λ integrase, generating attL and attR junctions. The kanamycin resistance marker can subsequently be excised through FLP/FRT-mediated recombination using pCP20. (b) Transformation plate obtained from a single λ integrase-mediated integration experiment after introducing approximately 0.02 pmol of the circularized donor ligation product into 50 μL of *E. coli* MG1655 competent cells. More than 10^4 antibiotic-resistant colonies were observed under the complete integration protocol. (c) Colony PCR analysis of 16 colonies picked from the plate shown in b. All 16 tested colonies yielded band patterns consistent with the expected integration junctions. Primer pairs F1/R2 and F2/R1 were used to detect the left and right integration junctions, respectively, whereas F1/R1 was used to distinguish the wild-type attB locus from the integrated locus.

In *S. cerevisiae*, the genomic integration module supports both classical homologous recombination and CRISPR-Cas9-assisted integration [3,30,31]. Level 3 integration backbones contain locus-specific homology arms and selectable markers, allowing assembled constructs to enter predefined genomic backgrounds directly. When greater site specificity or integration efficiency is required, the same class of constructs can additionally be combined with Cas9/sgRNA routes. The integration backbones provided by the toolkit cover multiple selectable markers and multiple chromosomal loci, and

reporter modules and transcription-factor modules can be fixed in different chromosomal contexts to provide stable genetic environments for subsequent part characterization and system testing (Supplementary Table 1, Supplementary Note 3). Both integration routes generate stable transformants that can be used for fluorescence verification and functional characterization (Fig. 3, Supplementary Fig. 2), thereby reducing the additional variation introduced by episomal plasmids in yeast part testing.

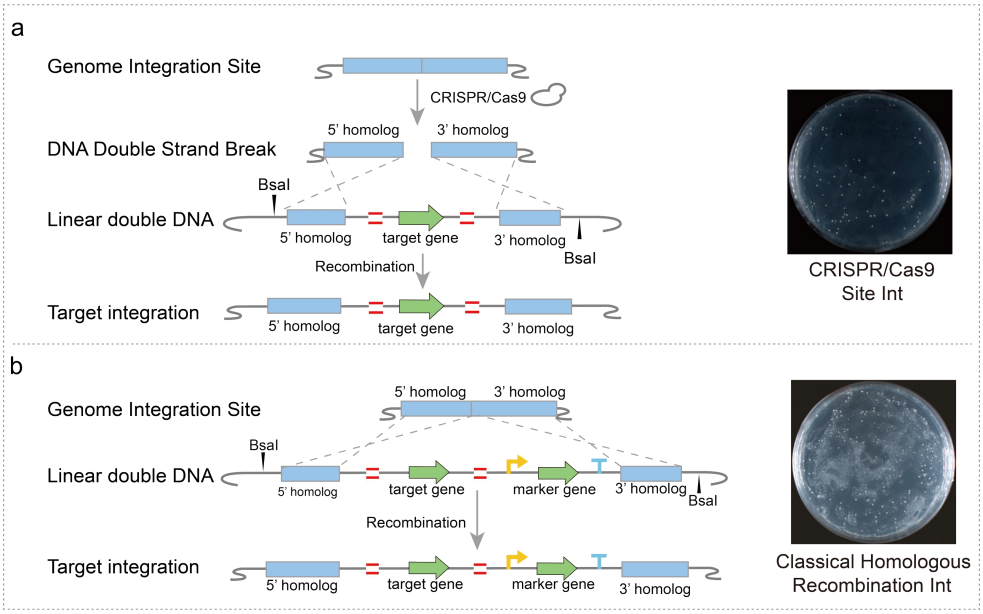


Fig. 3. Genomic integration in *S. cerevisiae*. (a) Schematic diagram of the Cas9-based recombination strategy. The Cas9 protein carrying the sgRNA introduces a double-strand break at the genomic locus on chromosome X (X: 236336...237310). Homologous recombination then occurs between the homology arms flanking the exogenous linearized DNA and the corresponding genomic locus, enabling specific integration of the target fragment. (b) Conventional homologous recombination-based integration in yeast. Homologous recombination occurs between the homology arms flanking the exogenous linearized DNA and the IIA locus on chromosome V, integrating the target fragment together with the MAR4_CaURA3 marker gene into the corresponding site, thereby achieving targeted integration at a conventional genomic locus.

Thus, in this toolkit, Level 3 is not simply the endpoint of multi-transcription-unit assembly, but also the switching point through which the same construct can enter different stable genomic backgrounds. Although *E. coli* and *S. cerevisiae* rely on different host-specific integration mechanisms, both follow the same logic in the present framework: constructs are first assembled under a unified set of rules and are then transferred into defined testing backgrounds by switching to appropriate Level 3 integration backbones,

thereby providing a stable basis for subsequent standardized quantitative characterization.

3.3. Standardized testing frameworks in the two hosts

After establishing hierarchical construction and genomic integration schemes, we further implemented unified frameworks for part and system testing in both hosts. The relevant test objects are not isolated sequences, but performance objects defined under fixed backbones, fixed host backgrounds, fixed reference parts, and fixed readout conditions. On this basis, constructs generated with the toolkit can enter standardized quantitative characterization pipelines directly, without the need to redesign testing logic for each new part or regulatory system.

In *E. coli*, promoter, ribosome binding site, and terminator characterization all use a unified testing architecture (Fig. 4a, d, f). All part-testing plasmids use a p15A replicon and gentamicin resistance in order to minimize variation from vector background and selection conditions. Promoter activity was quantified relative to BBa_J23101, and representative promoters including BBa_J23119, BBa_J23101, BBa_J23107, and others were characterized in relative promoter units (RPUs) [12] (Fig. 4a, b). Ribosome binding sites were quantified relative to BCD14c, and BCD2c, BCD14c, BCD21c, and others were compared in relative expression units (REUs) [13] (Fig. 4a, c). Terminators were evaluated in an upstream/downstream dual-reporter framework [14,26]. Using DT101 as the reference, the relative strengths of DT101, DT100, and DT5 were determined [26] (Fig. 4d, e; Supplementary Fig. 3). For transcription-factor systems, we further separated testing into two objects: an input module carried on a low-copy pSC101 backbone with kanamycin resistance for stable transcription-factor expression, and an output module carried on p15A with gentamicin resistance to read downstream promoter responses. This arrangement makes inducer concentration, transcription-factor level, and output signal quantifiable in a stepwise manner rather than conflating them within a single construct.

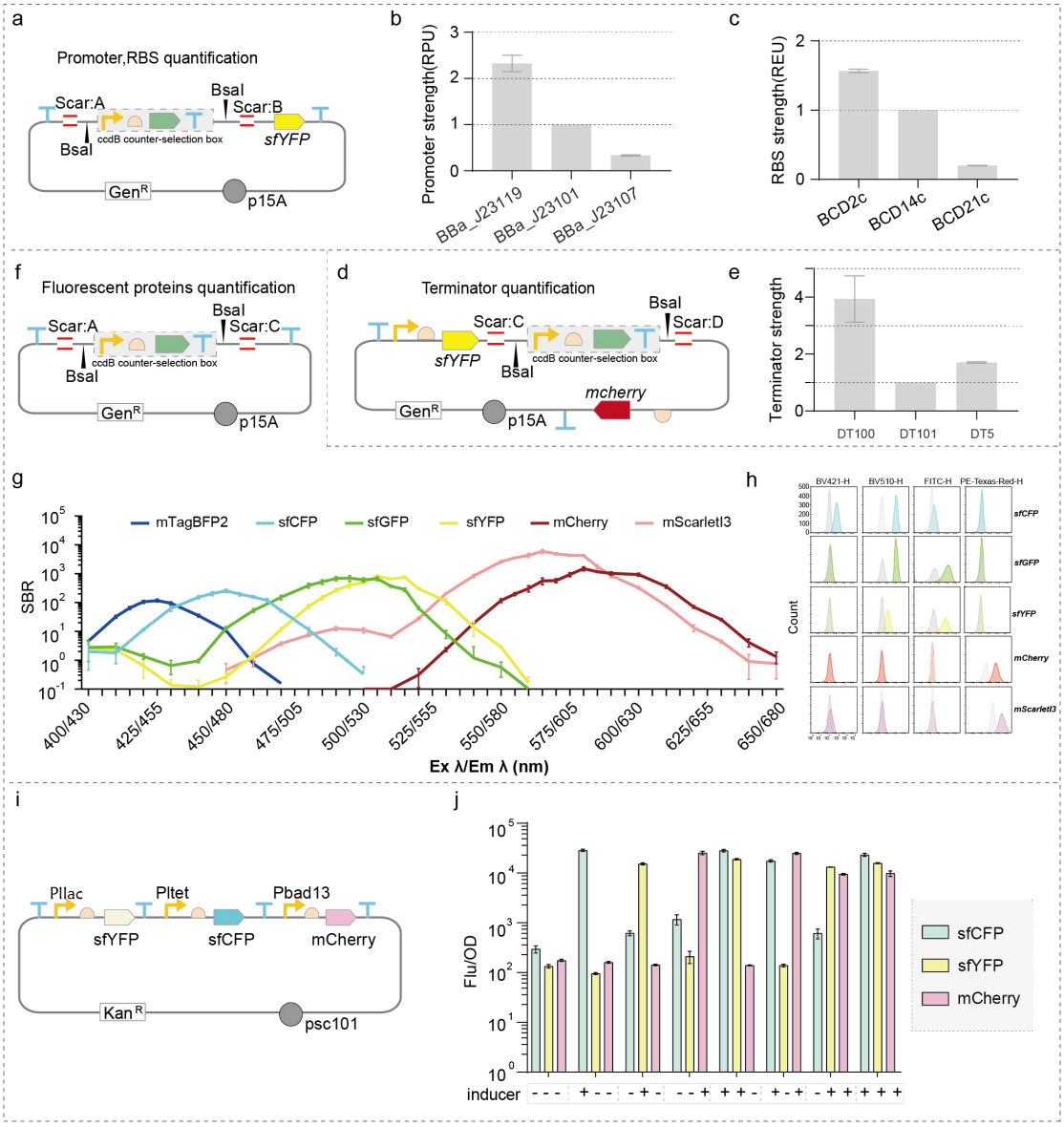


Fig. 4. Design of test backbone plasmids and characterization of genetic parts in *E. coli*. (a) Test backbone plasmid for promoter and RBS characterization. This backbone was designed in the A B scar format and requires simultaneous assembly of a promoter and an RBS. (b) Titration of promoter strength. The relative promoter units (RPUs) of each promoter were calculated using BBa_J23101 as the reference part. (c) Titration of RBS strength. The relative expression units (REUs) of each RBS were calculated using BCD14c as the reference part. (d) Test backbone plasmid for terminator characterization. This backbone contains a dual-fluorescent reporter system, in which upstream sfYFP is induced while downstream mCherry readthrough is simultaneously measured. (e) Titration of terminator strength. The relative strength of each terminator was calculated using DT101 as the reference part. (f) Test backbone plasmid for fluorescent protein characterization. This backbone was designed in the AC scar format and requires simultaneous assembly of a promoter, an RBS, and a CDS. (g) Spectral scan results of fluorescent proteins. Excitation scans were performed with a step size of either 5 or 10 nm, while the wavelength interval between excitation and detected emission was fixed at 30 nm. E_X: excitation; E_M: emission. (h) Flow cytometry plots of sfCFP, sfYFP, sfGFP, mCherry, and mScarletI3 in the BV421-H, BV510-H, FITC-H and PE-Texas Red-H channels. No compensation was applied. The gray curves

represent the uninduced condition, whereas the curves shown in the corresponding fluorescence colors represent the induced condition. (i) Schematic diagram of the multicolor co-expression system in cells. (j) Test results of the multicolor expression system. sfCFP, sfYFP, and mCherry were induced by aTc, IPTG, and arabinose, respectively. The x-axis indicates the presence (+) or absence (-) of the inducers in the order of aTc, IPTG, and arabinose. The signals were quantified as Flu/OD_{600} using a microplate reader (see Materials and methods). Data are shown as mean ± SD, with n = 3 independent replicates on different days per experiment.

In *S. cerevisiae*, the standardized testing framework places even greater emphasis on stable genetic background. Part characterization of promoters and terminators is performed primarily in predefined integration backgrounds: the reporter module is integrated at the IIA locus on chromosome V using MAR4_CaURA3 as the selection marker [32], whereas transcription-factor modules are integrated at the IIB locus on chromosome III or at the I locus on chromosome XV using MAR2_KlLEU2 or MAR7_clonNAT [33], respectively. Promoter testing uses the pScbb23 backbone, in which only the promoter is replaced (Fig. 5a, b), and pScLv305 (SpacerD1-pPFY1-mCitrine-tENO1) serves as the reference construct. Terminator testing uses pScbb46 and pScbb47, in which the terminator is exchanged within a fixed expression cassette to compare its effect on expression output (Fig. 5c, d; Supplementary Fig. 4c, d). Similar to the input/output separation used in *E. coli*, system testing in yeast is also organized into separate transcription-factor expression modules and reporter modules. In this way, constitutive promoters, inducible promoters, terminators, and transcription-factor systems can all be brought into a common comparative framework (Fig. 5a, c, e; Supplementary Fig. 4a, b, c).

To reduce uncertainty at the readout level, we also calibrated recommended detection windows and channel interference for representative fluorescent reporters in both hosts. In *E. coli*, excitation scans of mTagBFP2, sfCFP, sfGFP, sfYFP, mCherry, and mScarletI3 defined recommended detection wavelengths and showed substantial spectral overlap between sfGFP and sfYFP as well as between mCherry and mScarletI3 (Fig. 4f, g; Supplementary Fig. 5a-e). In *S. cerevisiae*, scans and flow-cytometric analysis of mCitrine, mNeonGreen, mScarletI3, mRuby2, and mTagBFP2 likewise defined recommended detection channels and indicated measurable interactions between green/yellow and

some red fluorescent proteins (Fig. 5e-h; Supplementary Fig. 6). Thus, in this toolkit fluorescent proteins serve not only as reporters, but also as part of the standardized testing conditions that support unified detection windows for subsequent multicolor readout and system testing.

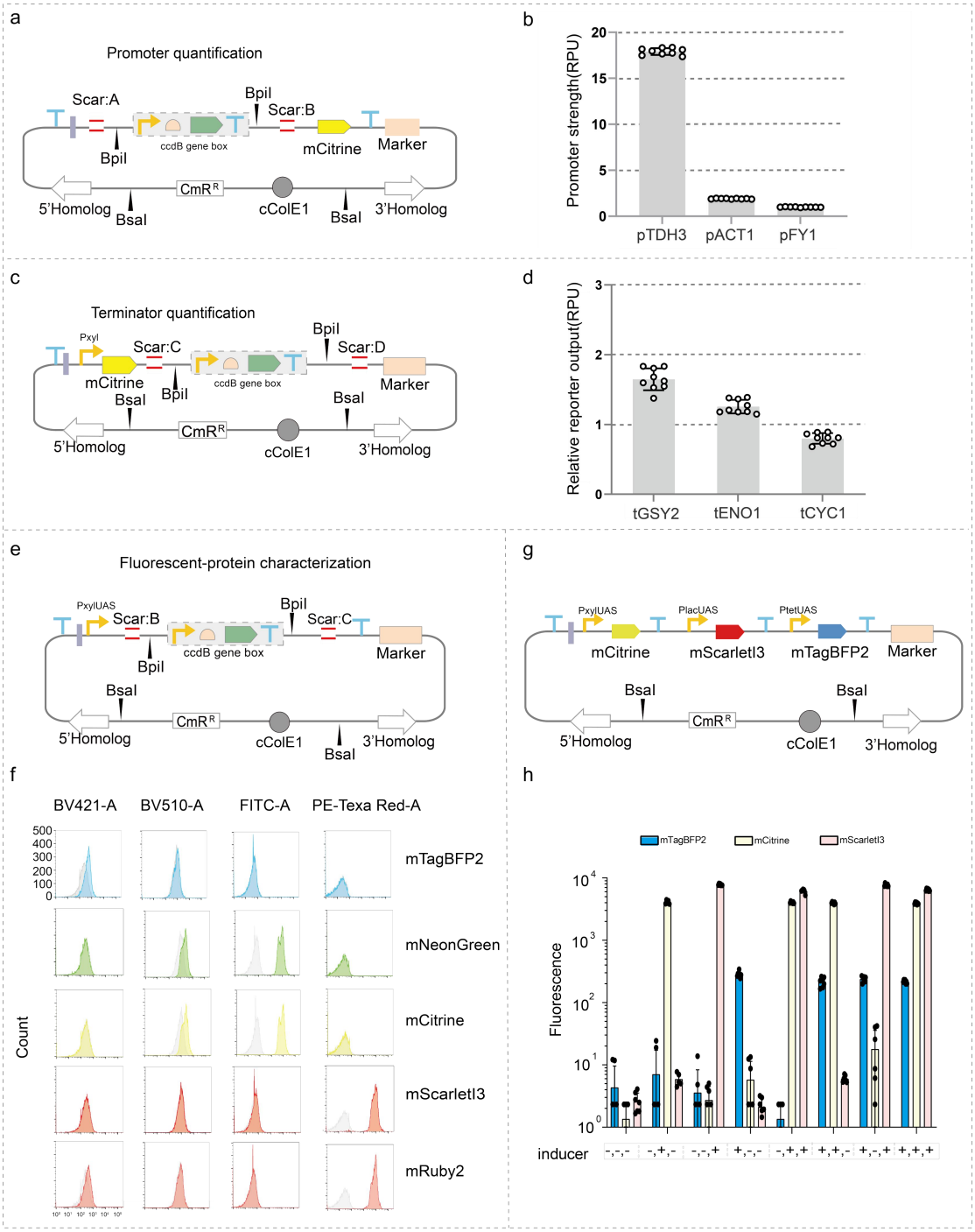


Fig. 5. Characterization of genetic parts in yeast. (a) Test backbone plasmid for promoter characterization. (b) Results of promoter strength characterization. The relative promoter units (RPUs) of individual promoters were calculated using the strain carrying the integrated pScLv305 reference construct as the reference. (c) Test backbone plasmid for terminator characterization. (d) Results of terminator

characterization using relative reporter output (RPU). Relative terminator activities were calculated using the strain carrying the integrated pScLv305 construct as the reference. (e) Test backbone plasmid for fluorescent protein spectral scanning. (f) Flow cytometric signal distributions of mTagBFP2, mNeonGreen, mCitrine, mScarletI3, and mRuby2 in the BV421-A, BV510-A, FITC-A, and PE-Texas Red-A channels. The gray curves represent the uninduced condition, whereas the curves shown in the corresponding fluorescence colors represent the induced condition. (g) Schematic diagram of the multicolor co-expression system in cells. (h) Test results of the multicolor co-expression system. Fluorescence signals were measured for eight combinations of three inducers, and the results were quantified using the median signal intensities from flow cytometry in the BV421-A, FITC-A, and PE-Texas Red-A channels. mCitrine, mScarletI3 and mTagBFP2 were induced by xylose (20 mmol/L), IPTG (20 mmol/L) and aTc (100 ng/mL) respectively. Data are shown as mean ± SD, with n = 3 independent replicates on different days per experiment.

Taken together, the toolkit defines not only standardized assembly rules but also standardized testing rules. In *E. coli*, comparable test objects are built through fixed vector backgrounds, fixed references, and a separated input/output logic. In *S. cerevisiae*, stable quantitative characterization is enabled through fixed integration backgrounds, fixed reference architectures, and fixed readout conditions. This dual-host testing framework creates a reusable experimental entry point for future part accumulation, system comparison, and cross-condition calibration.

3.4. Representative regulatory systems validate the standardized workflow

To test whether the standardized construction, genomic integration, and testing framework can support functional objects in real cellular contexts, we assembled and characterized representative regulatory systems in both hosts. The purpose here was not to maximize circuit complexity, but to examine whether, under predefined backbones, host backgrounds, and readout conditions, the toolkit can reliably organize standardized parts into functional modules with interpretable input-output relationships.

In *E. coli*, we constructed a LasR activation system and a LacI repression system. In both cases, the design was separated into an input module, which defines transcription-factor expression, and an output module, which reads the fluorescence response of the target promoter. Input titration based on the PcymRC system first established the operating window for transcription-factor expression, with the curve reaching saturation under high concentrations of cuminic acid (CA). Ligand-response

curves were then measured under selected input conditions (Supplementary Fig. 7a). The LasR system showed the expected activation response to N-(3-oxododecanoyl)-L-homoserine lactone (3O-C12-HSL) (Fig. 6a; Supplementary Fig. 7b), and the LacI system showed signal recovery upon IPTG-mediated derepression (Fig. 6b; Supplementary Fig. 7c). Across different input levels, both systems retained interpretable response trends, indicating that the framework can separate input definition from output readout and support quantitative testing under defined conditions.

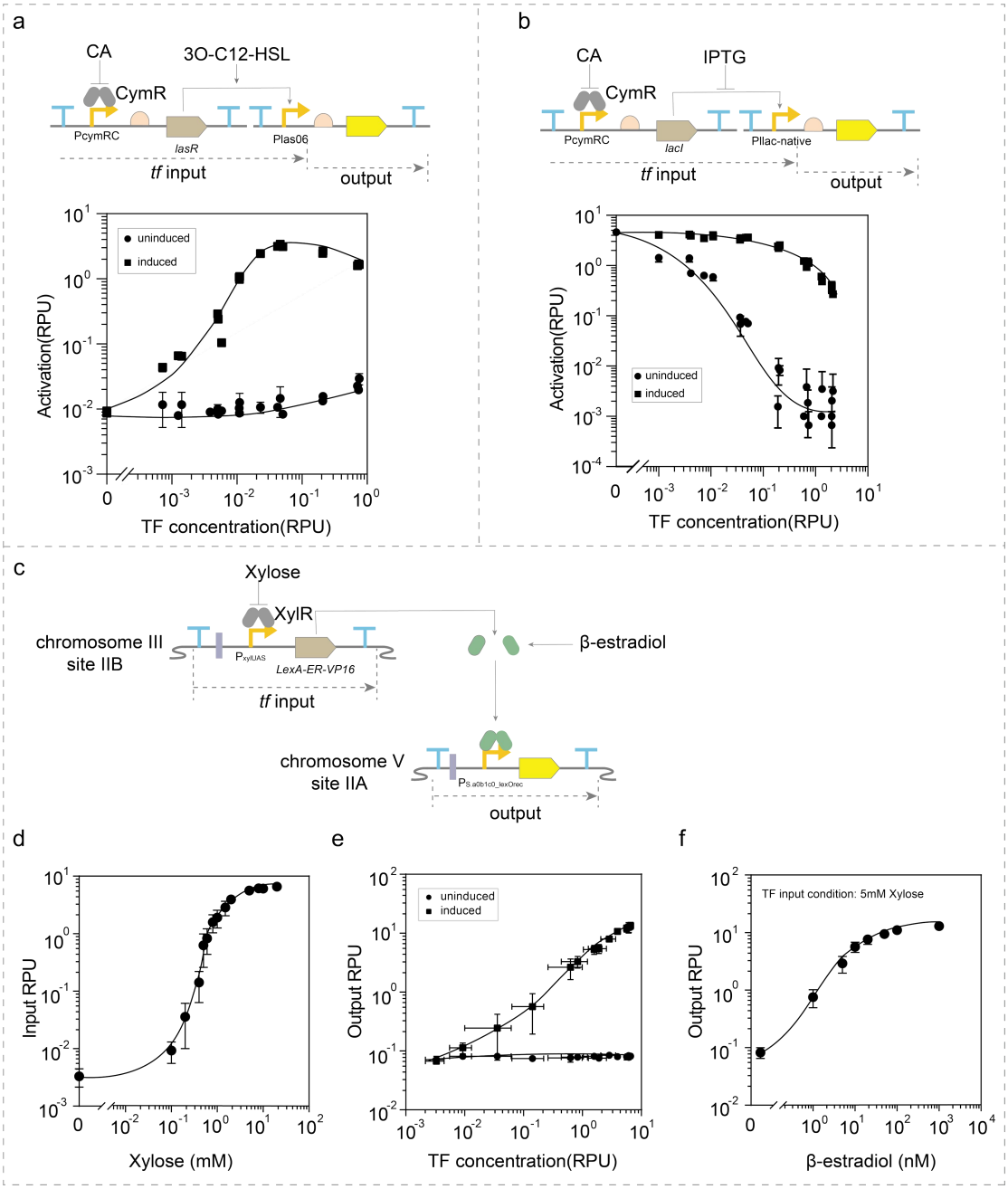


Fig. 6. Characterization of gene circuits in *E. coli* and *S. cerevisiae*. (a) Schematic diagram and system output of the LasR-activated gene circuit in *E. coli*. The x-axis indicates transcription factor (TF) expression level (RPU), regulated by the PcymRC-CymR system. The y-axis indicates reporter output (RPU), showing the output strength at different TF concentration gradients in the absence of ligand or in the presence of 1 μM 3O-C12-HSL. (b) Schematic diagram and system output of the LacI-repressed gene circuit in *E. coli*. The x-axis is the same as in a, and the y-axis shows reporter output (RPU), indicating the output strength at different TF concentration gradients in the absence of ligand or in the presence of 1 mM IPTG. For the LacI system, 100 μM cuminic acid (CA) was added during overnight culture. (c) Schematic diagram of the ER-activated gene circuit in *S. cerevisiae*. (d) TF input titration curve measured with CYE72 strain carrying the integrated pScLv013 construct. TF concentration was represented by the activity of a matched promoter-reporter construct rather than by direct measurement of TF protein concentration. (e) System reporter output at different TF concentrations in the absence and presence of the ligand molecule 1 μM β-estradiol, and (f) the induction curve of the ER system in response to β-estradiol at the optimal TF expression level (input: xylose, 5 mM). Data are shown as mean ± SD, with n = 3 independent replicates on different days per experiment.

In *S. cerevisiae*, we further assembled and tested representative regulatory systems including XylR-PxylUAS, XylR/XlnR-Pxln.2b-xyl, and LacI-PlacUAS (Supplementary Fig. 8a-c). Under unified integration backgrounds and readout conditions, these systems all produced output changes in the expected direction in response to their cognate inducers, indicating that the toolkit supports standardized construction and testing of both activation-type and repression-type regulatory objects in a eukaryotic host. In addition, a modular LexA-ER-VP16 system yielded interpretable transcription-factor input curves and β-estradiol response curves (Fig. 6c-f), showing that input modules, signal-conversion modules, and output modules can also be continuously defined within the same framework in yeast.

These results demonstrate that the dual-host standardized workflow established here is sufficient to support the full progression from assembly to testing for representative regulatory systems. More importantly, interpretation of these results does not depend on ad hoc replacement of vectors, hosts, or assay conditions, but instead rests on the unified backgrounds, references, and readout logic defined above. What is therefore validated here is not simply an individual system, but a reusable standardized entry point for construction and quantitative testing that operates under the same logic in both prokaryotic and eukaryotic chassis and provides a direct basis for future part accumulation, system comparison, and cross-condition calibration.

3.5. An automated plasmid assembly platform supports execution of the hierarchical workflow

The hierarchical definitions and interface rules of the toolkit are suitable not only for manual cloning but also for digital implementation. Based on the same Level 1–3 object definitions, scar rules, and backbone-switching relationships, we developed a companion automated plasmid assembly platform for plasmid-map visualization, hierarchical path generation, and assembly management. At present, the platform primarily supports digital execution from standard part retrieval to multi-level construct generation and does not directly undertake logical circuit design itself [8,9]. Because part types, level transitions, and backbone choices are predefined within the toolkit, the platform can generate assembly routes directly from selected parts and target constructs, allowing the same standardized rules to operate in both experimental execution and digital management.

To validate compatibility between the platform and the toolkit hierarchy, we used the yeast system as an example and selected six Level 1 part plasmids. The platform automatically combined them into two Level 2 transcription units and then generated one Level 3 multi-transcription-unit plasmid (Fig. 7). This example demonstrates that the hierarchical design, interface definitions, and backbone-switching relationships are sufficiently explicit to be encoded as executable digital rules, thereby supporting continuous progression from standard part retrieval to multi-level construct generation. The deployed platform and source code are available as described in the Data availability section.

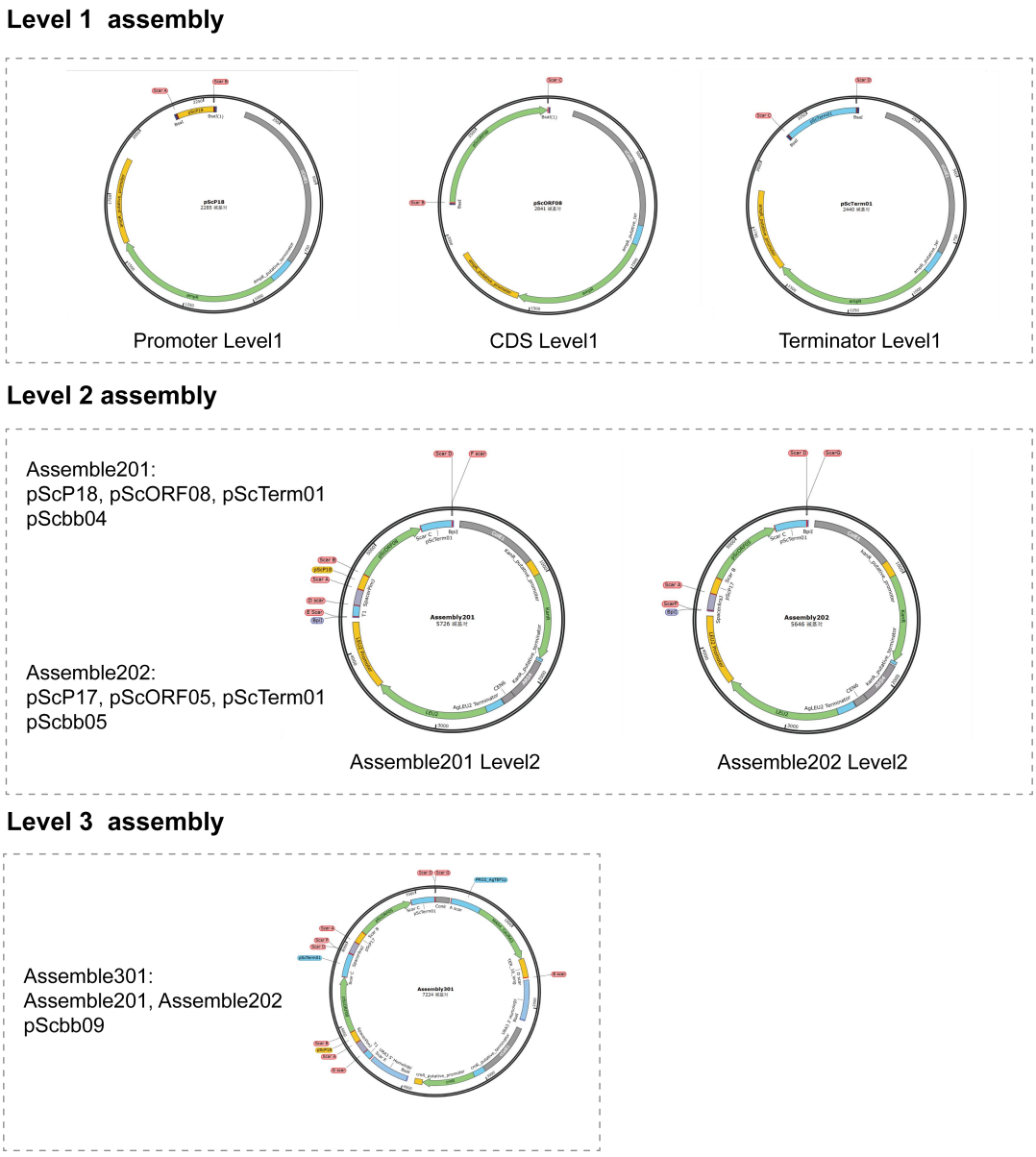


Fig. 7. Automation-platform-mediated multilevel plasmid assembly workflow. Starting from six Level 1 part inputs comprising five unique part plasmids from the *Saccharomyces cerevisiae* toolkit, which cover three types of functional elements—promoter, CDS, and terminator—each promoter–CDS–terminator set is assembled into a Level 2 plasmid carrying a single transcriptional unit. Subsequently, two Level 2 plasmids are further assembled into one Level 3 plasmid containing multiple transcriptional units. This workflow demonstrates the sequential assembly capability of the automation platform for standardized part selection, single transcriptional unit construction, and multigene module combination.

The significance of the platform in this study is therefore not that it replaces genetic design itself, but that it demonstrates that the rules of this dual-host toolkit are explicit enough to be translated into a stable and executable digital assembly workflow. In this way, it further increases the value of the toolkit for standardized part management, hierarchical construction, and future integration with automation.

4. Discussion

The Polaris toolkit was developed not simply as a collection of plasmid backbones, but as an execution framework that connects genetic-part storage, hierarchical construction, genomic implementation, and quantitative characterization. Its practical value lies in organizing several transitions that are frequently treated as separate tasks: how basic parts enter a construction system, how they are upgraded into transcription units and multigene constructs, how completed constructs are transferred into defined genetic backgrounds, and how the resulting objects are quantitatively tested. By predefining these transitions, Polaris is intended to reduce the need for users to independently coordinate backbone selection, assembly level, screening strategy, integration route, and testing architecture before routine experiments can begin.

This execution framework arises from the integration of several complementary design choices rather than from any single assembly feature. The Level 1–3 hierarchy assigns explicit identities and upgrade paths to reusable parts, intermediate transcription units, and final application constructs. The separation of part-level and transcription-unit-level interfaces distinguishes genetic-part boundaries from higher-level module boundaries, while the hierarchical use of BpiI and BsaI assigns clear roles to different assembly stages. Parallel ccdB- and mScarletI3-based backbone versions provide alternative screening strategies for laboratories with different propagation strains and screening requirements. Level 3 backbones further serve as switching points through which assembled constructs can enter plasmid-based testing configurations or host-specific genomic-integration routes. Importantly, the *E. coli* and *S. cerevisiae* implementations share the same object hierarchy and assembly progression while retaining distinct part architectures, backbone designs, integration mechanisms, and testing procedures appropriate for each host.

The standardized testing framework is another central component of Polaris. In this context, a genetic part is treated not only as a DNA sequence, but as a performance object defined by its assembly interfaces, vector context, host background, reference construct, cultivation condition, and measurement procedure. Fixed testing architectures

and matched reference constructs allow promoter, ribosome-binding-site, terminator, fluorescent-protein, and regulatory-system measurements to be interpreted within defined experimental contexts. The separation of transcription-factor input from reporter output further enables regulatory systems to be described through stepwise input–output relationships rather than through inducer concentration alone. This organization provides a practical basis for accumulating additional parts and regulatory modules under a common testing logic.

At the same time, the quantitative values generated using Polaris should be interpreted as context-dependent rather than universally transferable constants. Measurements obtained in the present study remain conditional on the specified host strain, vector or integration background, growth condition, reference construct, and measurement platform. Consequently, direct transfer of quantitative values between laboratories or experimental platforms will require appropriate reference measurements and recalibration. The contribution of the current toolkit is therefore to define the objects and conditions required for comparison, rather than to eliminate all sources of biological and experimental variation.

The explicit object definitions and transition rules also make Polaris suitable for digital implementation. The companion platform encodes part types, backbone choices, interface compatibility, and Level 1–3 transitions, allowing users to retrieve components, generate hierarchical assembly routes, and visualize resulting plasmid designs. Its current function is assembly planning and construct management rather than de novo circuit design or robotic execution of cloning experiments. Nevertheless, the deployed platform and publicly available source code provide an extensible basis for connecting physical toolkit resources with future databases, automated construct management, and machine-readable design–build–test workflows.

Several boundaries of the present implementation should be noted. First, the two host systems share a common hierarchical logic, but direct sequence-level interchangeability between *E. coli* and *S. cerevisiae* parts is not implied. Second, the Int2, Int5, and Int7 integration systems are included as expansion resources but were not experimentally validated in this study. Third, the demonstrated regulatory systems and

characterized parts represent practical examples of the workflow rather than an exhaustive survey of all possible components or circuit architectures. Finally, the digital platform currently supports hierarchical assembly planning and visualization but does not perform logical circuit design or automated experimental execution.

Future development can proceed along three interconnected directions. The first is the expansion of standard reference objects, including additional characterized parts, integration backbones, host backgrounds, and regulatory systems. The second is the continued development of a structured data layer that links sequences with genetic context, experimental conditions, reference measurements, and quantitative performance. The third is the extension of the digital platform toward more complete construct management, data return, and automation-compatible execution. Along this trajectory, Polaris can serve as a physical and digital entry point for the gradual accumulation of comparable genetic objects and reusable experimental workflows.

5. Conclusions

In conclusion, we established Polaris, a dual-host Golden Gate toolkit that connects hierarchical DNA assembly, host-specific genomic integration, and quantitative testing in *E. coli* and *S. cerevisiae*. The toolkit defines a shared three-level object hierarchy, distinguishes part-level from transcription-unit-level interfaces, and combines hierarchical BpiI/BsaI assembly with parallel ccdB- and mScarletI3-based screening options. Host-specific integration modules and defined testing architectures enable assembled constructs to proceed directly into part characterization and regulatory-system analysis, as demonstrated with representative genetic parts and activation and repression systems in both hosts. A companion digital platform further encodes the Level 1–3 rules for component selection, compatibility checking, plasmid visualization, and multilevel assembly planning. Rather than introducing a new assembly chemistry alone, Polaris provides an integrated execution framework that can support future part expansion, quantitative data accumulation, database development, and automation-compatible synthetic biology workflows.

CRediT authorship contribution statement

Ronghui Xie: Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft. Ran Wang: Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft. Yi Zhan: Investigation, Methodology, Resources, Validation. Boyan Wei: Methodology, Resources, Software. Ye Chen: Conceptualization, Funding acquisition, Formal analysis, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. Lili Liu: Conceptualization, Resources, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA0510200). We thank former and current members of our laboratory for providing the genetic parts.

Appendix A. Supplementary data

Supplementary material associated with this article is provided as separate files, including Supplementary Information, plasmid and strain inventories, genetic part sequences, and quantitative characterization datasets for *Escherichia coli* and *Saccharomyces cerevisiae*.

Data availability

The Polaris Kit Assembly source code and all 240 plasmid maps (Kits_plasmids_map.zip) are available at https://github.com/ChenYe-Lab/PolarisKitAssembly. Processed quantitative data, part sequences, and plasmid inventories are provided in the Supplementary Materials. The deployed web platform is available at https://infrasynbio.siat.ac.cn/kitassembly.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) for English-language editing. The authors reviewed the text and take full responsibility for the content of the manuscript.

References

[1] Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. PLoS One 2008;3:e3647. https://doi.org/10.1371/journal.pone.0003647.

[2] Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. A modular cloning system for standardized assembly of multigene constructs. PLoS One 2011;6:e16765. https://doi.org/10.1371/journal.pone.0016765.

[3] Lee ME, DeLoache WC, Cervantes B, Dueber JE. A highly characterized yeast toolkit for modular, multipart assembly. ACS Synth Biol 2015;4:975-86. https://doi.org/10.1021/sb500366v.

[4] Iverson SV, Haddock TL, Beal J, Densmore DM. CIDAR MoClo: improved MoClo assembly standard and new *E. coli* part library enable rapid combinatorial design for synthetic and traditional biology. ACS Synth Biol 2016;5:99-103. https://doi.org/10.1021/acssynbio.5b00124.

[5] Moore SJ, Lai HE, Kelwick RJR, Chee SM, Bell DJ, Polizzi KM, et al. EcoFlex: a multifunctional MoClo kit for *E. coli* synthetic biology. ACS Synth Biol 2016;5:1059-69. https://doi.org/10.1021/acssynbio.6b00031.

[6] Blázquez B, San León D, Torres-Bacete J, Gómez-Luengo A, Kniewel R, Martínez I, et al. Golden Standard: a complete standard, portable, and interoperative MoClo tool for model and non-model proteobacteria. Nucleic Acids Res 2023;51:e98. https://doi.org/10.1093/nar/gkad758.

[7] Mejlsted J, Kubaczka E, Wirth S, Koeppl H. The Coli Toolkit (CTK): an extension of the modular yeast toolkit for use in *E. coli*. ACS Synth Biol 2026;15:462-71. https://doi.org/10.1021/acssynbio.5c00489.

[8] Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, et al. Genetic circuit design automation. Science 2016;352:aac7341. https://doi.org/10.1126/science.aac7341.

[9] Chen Y, Zhang S, Young EM, Jones TS, Densmore D, Voigt CA. Genetic circuit design automation for yeast. Nat Microbiol 2020;5:1349-60. https://doi.org/10.1038/s41564-020-0757-2.

[10] Ham TS, Dmytriv Z, Plahar H, Chen J, Hillson NJ, Keasling JD. Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. Nucleic Acids Res 2012;40:e141. https://doi.org/10.1093/nar/gks531.

[11] McLaughlin JA, Beal J, Mısırlı G, Grünberg R, Bartley BA, Scott-Brown J, et al. The Synthetic Biology Open Language (SBOL) version 3: simplified data exchange for bioengineering. Front Bioeng Biotechnol 2020;8:1009. https://doi.org/10.3389/fbioe.2020.01009.

[12] Kelly JR, Rubin AJ, Davis JH, Ajo-Franklin CM, Cumbers J, Czar MJ, et al. Measuring the activity of BioBrick promoters using an *in vivo* reference standard. J Biol Eng 2009;3:4. https://doi.org/10.1186/1754-1611-3-4.

[13] Mutalik VK, Guimaraes JC, Cambray G, Lam C, Christoffersen MJ, Mai QA, et al. Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods 2013;10:354-60. https://doi.org/10.1038/nmeth.2404.

[14] Chen YJ, Liu P, Nielsen AAK, Brophy JAN, Clancy K, Peterson T, et al. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. Nat Methods 2013;10:659-64. https://doi.org/10.1038/nmeth.2515.

[15] Cardinale S, Arkin AP. Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems. Biotechnol J 2012;7:856-66. https://doi.org/10.1002/biot.201200085.

[16] Lou C, Stanton B, Chen YJ, Munsky B, Voigt CA. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. Nat Biotechnol 2012;30:1137-42. https://doi.org/10.1038/nbt.2401.

[17] Bayer CN, Rennig M, Ehrmann AK, Nørholm MHH. A standardized genome architecture for bacterial synthetic biology (SEGA). Nat Commun 2021;12:5876. https://doi.org/10.1038/s41467-021-26155-5.

[18] Lux MW, Strychalski EA, Vora GJ. Advancing reproducibility can ease the ‘hard truths’ of synthetic biology. Synth Biol (Oxf) 2023;8:ysad014. https://doi.org/10.1093/synbio/ysad014.

[19] Lund BA, Leiros HKS, Bjerga GK. A high-throughput, restriction-free cloning and screening strategy based on ccdB-gene replacement. Microb Cell Fact 2014;13:38. https://doi.org/10.1186/1475-2859-13-38.

[20] Wang H, Bian X, Xia L, Ding X, Müller R, Zhang Y, et al. Improved seamless mutagenesis by recombineering using ccdB for counterselection. Nucleic Acids Res 2014;42:e37. https://doi.org/10.1093/nar/gkt1339.

[21] Zhang X, Teng C, Lyu K, Lyu S, Fan Y. ‘Two in One’ cloning vector applied for blunt-end and T-A cloning with one-step digestion–ligation and screening of positive recombinants by unaided eyes. Curr Issues Mol Biol 2025;47:17. https://doi.org/10.3390/cimb47010017.

[22] Meyer AJ, Segall-Shapiro TH, Glassey E, Zhang J, Voigt CA. *Escherichia coli* “Marionette” strains with 12 highly optimized small-molecule sensors. Nat Chem Biol 2019;15:196-204. https://doi.org/10.1038/s41589-018-0168-3.

[23] Wang T, Xie R, Ma Z, Chen Y. A unified computational framework for quantitative design and optimization of transcriptional regulation across bacterial species. Nucleic Acids Res 2026;54:gkaf1472. https://doi.org/10.1093/nar/gkaf1472.

[24] Tafoya-Ramírez M, Padilla-Vaca F, Ramírez-Saldaña A, Mora-Garduño J, Rangel-Serrano A, Vargas-Maya N, et al. Replacing standard reporters from molecular cloning plasmids with chromoproteins for positive clone selection. Molecules 2018;23:1328. https://doi.org/10.3390/molecules23061328.

[25] Wu P, Li X, Yang M, Huang Z, Mo H, Li T, et al. High-throughput, one-step screening, cloning and expression based on the lethality of DpnI in *Escherichia coli*. Biochem Biophys Res Commun 2018;504:177-83. https://doi.org/10.1016/j.bbrc.2018.08.151.

[26] Park Y, Espah Borujeni A, Gorochowski TE, Shin J, Voigt CA. Precision design of stable genetic circuits carried in highly-insulated *E. coli* genomic landing pads. Mol Syst Biol 2020;16:e9584. https://doi.org/10.15252/msb.20209584.

[27] Yang L, Nielsen AAK, Fernandez-Rodriguez J, McClune CJ, Laub MT, Lu TK, et al. Permanent genetic memory with >1-byte capacity. Nat Methods 2014;11:1261-6. https://doi.org/10.1038/nmeth.3147.

[28] Landy A. Dynamic, structural, and regulatory aspects of λ site-specific recombination. Annu Rev Biochem 1989;58:913-41. https://doi.org/10.1146/annurev.bi.58.070189.004405.

[29] Li Q, Sun B, Chen J, Zhang Y, Jiang Y, Yang S. A modified pCas/pTargetF system for CRISPR-Cas9-assisted genome editing in *Escherichia coli*. Acta Biochim Biophys Sin (Shanghai) 2021;53:620-7. https://doi.org/10.1093/abbs/gmab036.

[30] Reider Apel A, d'Espaux L, Wehrs M, Sachs D, Li RA, Tong GJ, et al. A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. Nucleic Acids Res 2017;45:496-508. https://doi.org/10.1093/nar/gkw1023.

[31] Meng J, Qiu Y, Zhang Y, Zhao H, Shi S. CMI: CRISPR/Cas9 based efficient multiplexed integration in *Saccharomyces cerevisiae*. ACS Synth Biol 2023;12:1408-14. https://doi.org/10.1021/acssynbio.2c00591.

[32] Goldstein AL, Pan X, McCusker JH. Heterologous URA3MX cassettes for gene replacement in *Saccharomyces cerevisiae*. Yeast 1999;15:507-11. https://doi.org/10.1002/(SICI)1097-0061(199904)15:6<507::AID-YEA369>3.0.CO;2-P .

[33] Gueldener U, Heinisch J, Koehler GJ, Voss D, Hegemann JH. A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. Nucleic Acids Res 2002;30:e23. https://doi.org/10.1093/nar/30.6.e23.