PERSPECTIVES

TECHNOLOGIES AND TECHNIQUES — INNOVATION

Bricks and blueprints: methods and standards for DNA assembly

Arturo Casini, Marko Storch, Geoffrey S. Baldwin and Tom Ellis

Abstract | DNA assembly is a key part of constructing gene expression systems and even whole chromosomes. In the past decade, a plethora of powerful new DNA assembly methods — including Gibson Assembly, Golden Gate and ligase cycling reaction (LCR) — have been developed. In this Innovation article, we discuss these methods as well as standards such as the modular cloning (MoClo) system, GoldenBraid, modular overlap-directed assembly with linkers (MODAL) and PaperClip, which have been developed to facilitate a streamlined assembly workflow, to aid the exchange of material between research groups and to create modular reusable DNA parts.

Our ability to cut and paste DNA from different sources and to assemble it into gene constructs has been one of the key drivers of biological research and biotechnology over the past four decades. However, despite countless advances in molecular biology, the assembly of DNA parts into new constructs remains a craft that is both time consuming and unpredictable. The decreasing cost of gene synthesis promises to alleviate these limitations by providing custom-made double-stranded DNA fragments that are typically between 200 bp and 2,000 bp in length[1]. Nonetheless, gene synthesis does not eliminate the need for DNA assembly, which remains necessary for the production of constructs >1 kb, both in research laboratories and at gene synthesis companies. DNA assembly also enables projects with more complex experimental needs to be carried out. It is especially valuable for building diverse plasmid libraries and creating multicomponent systems, and has even been used to construct synthetic cells[2].

Addressing the limitations of DNA assembly methods has been one of the key goals of synthetic biology — a scientific discipline focused on the construction and testing of new or redesigned versions of genes, gene networks, pathways and cells[3,4]. In order to tackle projects of increasing scale and complexity, researchers have invested considerable efforts into developing new tools for DNA assembly, and matching them with improved, lower-cost gene synthesis (reviewed in REFS 1,5), as well as developing a suite of important new tools for genome editing (BOX 1). Owing to these combined advances, the field has reached the point at which even undergraduate students are able to construct entire eukaryotic chromosomes using gene synthesis and DNA assembly methods[6].

This acceleration in the scale of DNA assembly has enabled researchers to pursue complex construction projects, which instead require engineering approaches. In the past decade, important assembly methods such as Gibson Assembly and Golden Gate have been developed[7,8], which define new protocols for joining together DNA parts. Alongside these approaches, researchers have developed various physical standards such as modular overlap-directed assembly with linkers (MODAL)[9] and the modular cloning (MoClo) system[10] that define rules for the format of DNA parts that can be used with them. These physical standards facilitate the re-use of parts between experiments and the exchange of parts between research groups, and importantly provide modularity in construction. In this Innovation article, we focus on the recent advances in DNA assembly methods and describe their corresponding physical standards (summarized in TABLE 1).

Box 1 | **Genome editing**

Genome-editing technologies rewrite DNA sequences in a site-specific manner within cells and can be used alongside DNA assembly methods in the construction and engineering of genomes. Multiplex automatable genome engineering (MAGE) entails the delivery of synthetic DNA oligonucleotides into growing cells to mutate specific genomic sequences during DNA replication[70]. It works efficiently in *Escherichia coli* and was used impressively to recode all 321 TAG stop codons in the *E. coli* genome to TAA stop codons to provide a genomically recoded organism capable of utilizing non-standard amino acids[71]. Other genome-editing methods rely on homologous recombination (a technique known as recombineering); in *E. coli*, this can be used to insert or delete DNA at almost any desired locus through the λ-Red method, which utilizes a phage recombinase to recombine long regions of homology[72]. This has already enabled researchers to construct an *E. coli* genome of reduced size by performing large-scale deletions of unwanted elements[73]. In other organisms, the emergence of programmable nucleases such as zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and, most recently, the CRISPR–Cas9 (clustered regularly interspaced short palindromic repeat (CRISPR)–CRISPR-associated 9) genome-editing tool means that site-specific cutting of chromosomes is now also possible, paving the way for recombineering in these cells by homology-directed repair[74]. Genome-editing technologies differ from DNA assembly methods as they edit existing sequences rather than combine DNA parts together. Thus, combining the two approaches can be particularly powerful, as was recently demonstrated in the assembly and genomic integration of pathways of 14 genes into multiple loci of the yeast genome in one step by combining standardized DNA assembly methods with site-specific integration using CRISPR–Cas9 (REF. 75).

Table 1 | **Summary of physical standards in DNA assembly**

Physical standards	Underlying methodology					Limitations		Workflow		
	Restriction and ligation	HE	Type IIS RE	SSR	Long overlap	PCR required*	Forbidden restriction sites‡	Number of assembly tiers	Multipart assembly§	Hierarchical assembly
BioBrick[13] and BglBrick[15]	✓					No	4	1	No	Yes
iBrick[21]		✓				No	0	1	No	Yes
HVAS[22]		✓		✓		No	0	2	Yes; no	No‖
MoClo[29]			✓			No	3	2	Yes; yes	Yes
GoldenBraid 2.0 (REF. 30)			✓			No	3	≥2	Yes; no	Yes
GreenGate[32]			✓			No	1	2	Yes; no	Yes
Binder *et al.*[31]			✓			No	3	2	Yes; yes	Yes
PSA[37]			✓			No	0	1	No	Yes
DNA assembler[53]					✓	Yes	0	2	Yes; yes	No
MODAL[9]					✓	Yes	0	1	Yes	No
BASIC[58]			✓		✓	No	1	1	Yes	Yes
Torella *et al.*[55]¶	✓				✓	No	≥4#	2	No; yes	No
Guye *et al.*[59]		✓		✓	✓	No	0	2	Yes; yes	Yes
PaperClip[56]					✓	No	0	1	Yes	No

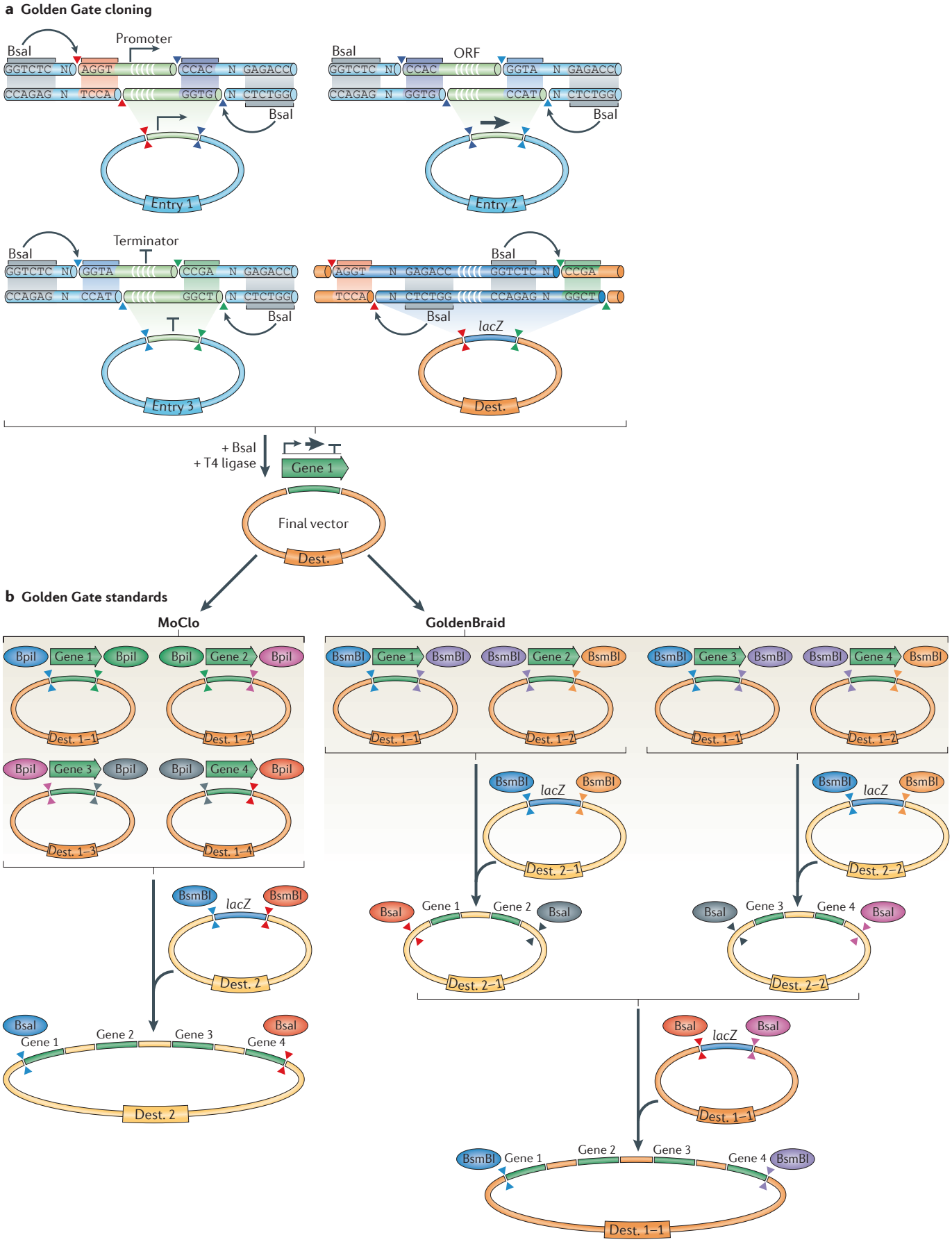
BASIC, biopart assembly standard for idempotent cloning; HE, homing endonuclease; HVAS, HomeRun Vector Assembly System; MoClo, modular cloning; MODAL, modular overlap-directed assembly with linkers; PSA, pairwise selection assembly; RE, restriction endonuclease; SSR, site-specific recombinase. *Single PCR amplifications that format DNA parts according to standard requirements (such as adding a prefix and a suffix) are not considered here. ‡The number of different restriction sites that need to be absent from all DNA parts. §Denotes whether the standard allows parallel assembly of more than two DNA parts at a time. If a standard uses two tiers, the first tier (left) and the second tier (right) are evaluated separately. GoldenBraid can use more than two tiers of assembly, but above tier one all assemblies are carried out two parts at a time. ‖The HVAS does not enable hierarchical rounds of assembly but enables some degree of post-assembly modification of the constructs. ¶The standard developed by Torella *et al.* in its latest version provides two options to attach linkers to parts. The table shows the option presented in the original paper. The alternative route uses a PCR amplification to attach the linkers, thereby removing any first-tier assembly[77]. #The BioBrick or BglBrick standard is used for first-tier assembly (with four forbidden sites) and a different digest is also required to begin second-tier assembly.

Endonuclease-mediated assembly

Gene cloning using restriction endonucleases and DNA ligases has been utilized successfully for more than 40 years[11], but in the age of cheap DNA synthesis its limitations are becoming evident. It typically only joins two DNA parts at a time and requires the parts of interest to be compatible with restriction sites located in a multiple cloning site within the target vector. The restriction sites chosen directly affect cloning efficiency and can even lead to substantial changes in plasmid function[10,12]. To resolve such experimental inconsistencies, standards for DNA assembly using restriction endonucleases began to emerge early this century. The BioBrick standard was one of the first and was developed to address the ad hoc nature of cloning[13]. BioBrick DNA parts are flanked by standard prefix and suffix sequences that contain four defined restriction sites. These sequences enable a repeatable, idempotent assembly process: the ligation of two BioBrick parts produces a new, larger BioBrick part with the same physical format. As the BioBrick standard became widely adopted, it was used as the basis to develop various new standards (for example, the BglBrick standard) with improved flexibility and efficiency[14–16].

Although standards such as BioBrick enable parts to be rationally assembled into desired constructs, it is often quicker to modify existing constructs. New standard plasmid formats have thus been developed to facilitate the swapping of parts between constructs. The breadboard standard has been developed for rapid prototyping of *Escherichia coli* gene network constructs[17]. Plasmids constructed using Standard European Vector Architecture (SEVA) can be used for working across a broad range of bacterial species. These plasmids have a modular structure and unique restriction sites flanking fundamental parts such as selection markers and origins of replication, which often need to be exchanged to function efficiently in different organisms[18,19].

The drawback of all digestion and ligation methods is the need to remove forbidden digestion sites within DNA parts before cloning them. Homing endonucleases such as I-SceI have been proposed as a way to overcome this, as they are equivalent to restriction endonucleases but only cut at long recognition sequences, which are unlikely to be found in cloning parts[20]. The iBrick standard closely resembles the BioBrick standard but uses prefix and suffix sequences containing homing endonuclease recognition sites[21]. The HomeRun Vector Assembly System (HVAS) also exploits homing endonucleases but within a tiered framework that enables multipart construction[22]. Although both iBrick and the HVAS tackle the issue of forbidden restriction sites, they result in the formation of large 'scar' sequences when parts are recombined as the long restriction endonuclease recognition sites remain in the final construct. Scar sequences can also pose a problem in BioBrick assembly, as the 8 bp sequence generated between two fused BioBricks can act as a destabilizing sequence in *E. coli*[23].



◂ Figure 1 | **Examples of type IIS restriction endonuclease-based methods and standards.** The Golden Gate method is a restriction endonuclease digestion-and-ligation-based technique that exploits the ability of the type IIS restriction endonuclease BsaI to cut DNA a few bases away from its recognition site, thereby leaving various overhangs that enable one-pot multipart assembly with high efficiency[24]. **a** | In an example of a first-tier Golden Gate assembly workflow, genes (green) are constructed from a series of genetic elements such as promoters, open reading frames (ORFs) and terminators. Multiple parts are cloned into 'entry vectors' that carry inward-facing BsaI restriction sites and are then assembled by a simultaneous digestion and ligation reaction into a 'destination vector ' (labelled dest.) that has outward-facing BsaI restriction sites and a drop-out screening cassette (for example, the *lacZ* gene). Digestion by BsaI generates different 4 bp overhangs (coloured bars) that define the position of the various parts in the final construct. Simultaneous digestion and ligation reactions force incorrectly assembled vectors to be recut by BsaI, whereas correctly assembled vectors are protected. **b** | The modular cloning (MoClo) system[29] and GoldenBraid 2.0 (REF. 30) standard provide different approaches for the second tier of Golden Gate assembly, in which the genes built in the first tier of assembly are combined to form multigene constructs. MoClo uses a parallel approach: in tier one, each gene is assembled into destination vectors flanked by BpiI digestion sites with different overhangs (shown as BpiI in different colours) and are digested in tier two to leave overhangs that define the final construct layout. GoldenBraid uses a pairwise approach: in tier one, genes are cloned into pairs of destination vectors (each containing the *lacZ* screening cassette), and these are then assembled two at a time in successive tiers by performing assembly alternately with BsmBI and BsaI type IIS restriction endonucleases.

Type IIS restriction endonuclease assembly. Type IIS restriction endonucleases differ from traditional type II restriction endonucleases because they cut DNA a few bases away from their recognition site, thereby providing the freedom to choose the overhang sequences they generate. The use of type IIS restriction endonucleases adds a substantial advantage to DNA assembly because it enables scar-free 'seamless' assembly, and just one enzyme can digest parts with different overhangs so that many of these can be ligated together in a defined order in one go. Golden Gate cloning is based on these principles and has gained notable popularity, especially in publicly available kits in which many parts and/or highly repetitive sequences need to be assembled in one step[24–26]. One of the main challenges of assembly using the Golden Gate method is defining the position of the DNA parts within the final construct, which depends on the sequences of the short overhangs generated by digestion.

Golden Gate parts can be generated by PCR or by gene synthesis[27,28], but to minimize errors in assembly the basic assembly parts are usually first placed in 'entry vectors' and then digested and ligated into 'destination vectors' (FIG. 1). To enable use of the Golden Gate method for hierarchical assembly, various physical standards have been developed that adopt a tiered assembly approach: in the first tier, genes are assembled from their elementary parts (for example, promoters, open reading frames and terminators)[29–32] (FIG. 1a); and, in the second tier, these genes are combined to form multigene systems (FIG. 1b). The MoClo standard, which was initially developed for plant molecular biology, uses a parallel approach for all tiers but requires a large number of entry and destination vectors[29]. The GoldenBraid standard reduces the number of required vectors by applying a pairwise approach for assembly beyond the gene level but at the cost of requiring more steps for larger constructs[30,33].

The ability of the Golden Gate method to perform parallel assembly of multiple parts without PCR has made it very popular, and recently a mammalian MoClo (mMoClo) standard was introduced[34]. The mMoClo standard enables the stable integration of very large multigene constructs into defined sites in mammalian chromosomes. Such constructs include genetic switch programmes that encode the regulated expression of counteracting transcription factors that establish hereditable memory in cells[34]. Golden Gate is best suited for applications in which multiple genes must be expressed together, and consequently it is also used in kits for multiplex genome editing by CRISPR–Cas9 (clustered regularly interspaced short palindromic repeat–CRISPR-associated 9)[35]. The large number of plasmids required for Golden Gate is a drawback, but in return researchers can format their DNA to physical standards that facilitate the re-use and sharing of parts. However, as with classic restriction endonuclease-based cloning, Golden Gate requires restriction endonuclease recognition sequences to be absent from the DNA parts to enable their cloning. To overcome this limitation, programmed DNA methylation (which can prevent digestion by certain endonucleases) has been utilized in some standards, such as GreenGate and the methylation-assisted tailorable ends rational (MASTER) ligation method, to guide digestion only to desired sites[32,36]. Impressively, methylation was also used in the related pairwise selection assembly method to construct a complete 91 kb synthetic yeast chromosome arm without the need to consider forbidden sites[37].

Site-specific recombination

Site-specific recombination omits any need for restriction endonucleases and instead uses phage integrases, which are site-specific recombinases that recognize versions of attachment (*att*) sequence motifs and catalyse DNA rearrangement between them. Integrases are utilized in the popular commercial Gateway cloning method[38], one of the earliest DNA assembly standards, which uses λ integrase *in vitro* to catalyse directional cloning of DNA parts that are flanked by orthogonal versions of the *attB* and *attP* sites recognized by the integrase (FIG. 2a). This method is simple, efficient and reliable, and is widely used for the generation of clone libraries and for expression analysis in eukaryotic systems[39,40]. By synthetically generating four additional orthogonal *att* recombination sequences, Gateway has also recently been expanded to enable the cloning of multiple parts simultaneously[41]. Similar non-commercial systems have also been developed that use alternative phage integrases, including the φBT1 and φC31 integrases[42,43]. For all of these methods, reactions at recombined *att* sites can be reversed by additionally providing either an excisionase (a bacteriophage excision protein) or a recombination directionality factor, which is an accessory protein that, in combination with the integrase, reverses the reaction and leads to excision rather than integration. For φC31 integrase-based assembly, excision enables the quick replacement of a single part within an already assembled construct, facilitating the insertion of an alternative part or of a multipart fragment that further expands the construct[43].

Owing to the nature of integrase sites, recombinase-based assembly leaves repeated scar sequences between all of the assembled parts, and these can be problematic for maintaining DNA integrity or for mRNA folding as they are typically long and palindromic. Furthermore, creating the entry vectors for recombinase-based cloning is not trivial, as the various *att* sites must each be matched with their specific partner sequence while also maintaining the correct orientation on the DNA strand (FIG. 2b).

Long-overlap-based assembly

Long-overlap-based assembly methods are *in vitro* and *in vivo* techniques that work by joining DNA fragments that share homologous sequences — usually between 20 bp and 40 bp in length — at their ends (FIG. 3a). The long sequence homology ensures high efficiency and specificity of DNA assembly, meaning that most long-overlap-based methods can easily assemble five or more DNA parts together in one step[7]. These methods are particularly popular for manipulating larger DNA fragments (>1 kb), as there is usually no requirement for the removal of restriction sites from within parts. Another reason for their popularity is that overlap regions can easily be added by PCR. The mechanism of action of these methods varies greatly: circular polymerase extension cloning (CPEC) is based on overlap extension PCR (OE-PCR[44]) and is essentially a high-fidelity PCR amplification in which the template and primers are replaced by the DNA fragments to be assembled into a plasmid[45]. As these are designed to share homology at their ends, the parts anneal to each other during PCR, and each part functions as a primer for the amplification of the other until eventually a nicked circular molecule is generated. Other methods use enzymatic digestions to convert the homologous regions of different parts into single-stranded overhangs that can directly anneal to each other. Uracil-specific excision reagent (USER) assembly does so by producing nicks in the homology regions[46,47], whereas Gibson Assembly uses a 'chew-back' mechanism on one strand, improving on the previously described sequence- and ligation-independent cloning (SLIC) method by the addition of *in vitro* DNA repair[2,48]. In these methods, purified repair enzymes can also be replaced with cellular DNA repair systems; for example, the seamless ligation cloning extract (SLiCE) method can assemble DNA parts *in vitro* at very low cost by using *E. coli* cell extracts instead of an enzyme cocktail such as that used by Gibson Assembly[49]. Other methods take this a step further by relying on the repair machinery of live organisms such as *Bacillus subtilis* and *Saccharomyces cerevisiae*, which are able to take up linear DNA parts with overlaps and assemble them spontaneously *in vivo* into the desired constructs by native homologous recombination[50–52].

Long-overlap assembly methods have very few sequence restrictions, but design requirements for the overlap sequences need to be considered if assembly is to be efficient, if modularity is required and

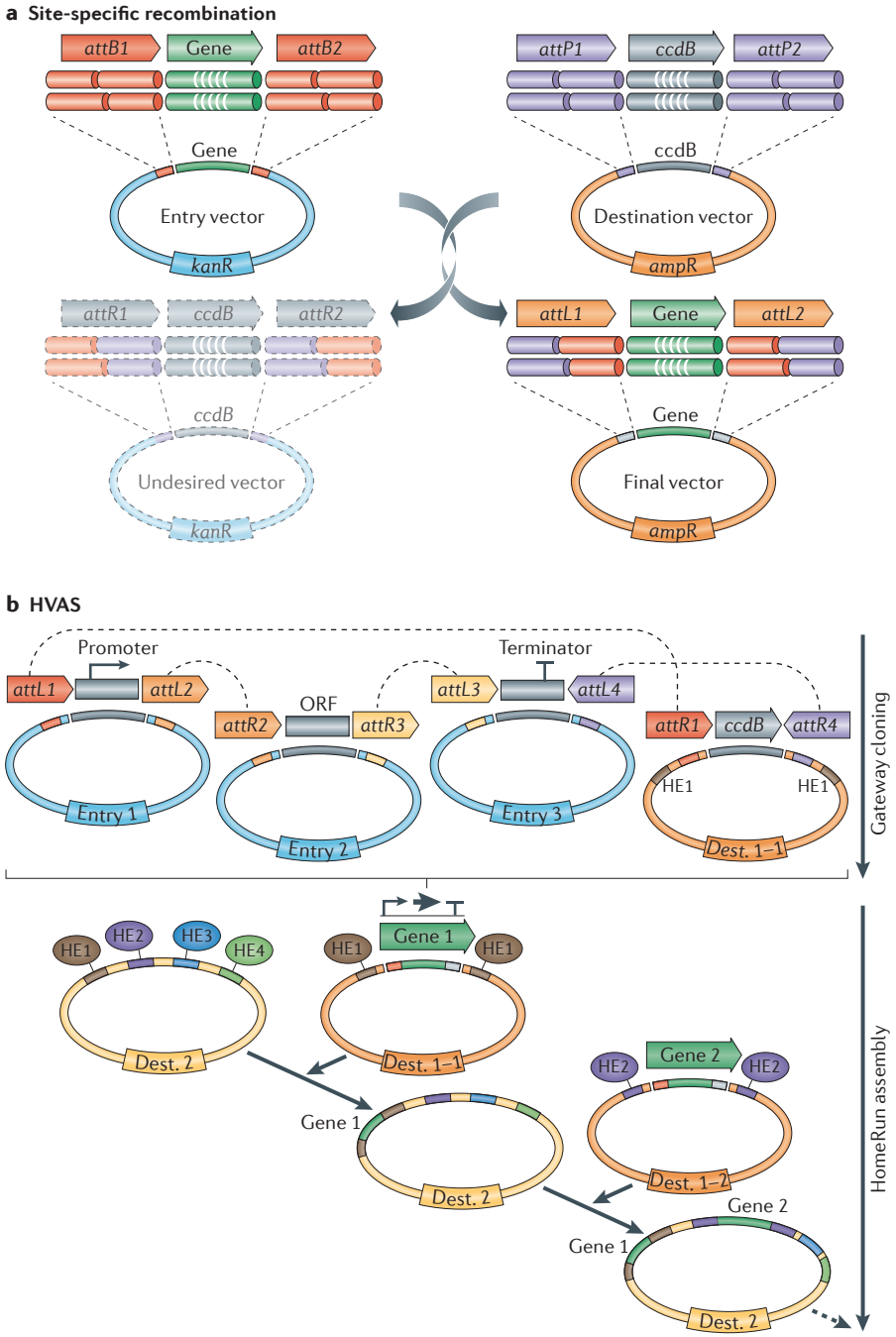


Figure 2 | **Site-specific recombination methods and standards.** **a** | A general diagram for site-specific recombination-based cloning is shown[76]. An 'entry vector' containing a gene flanked by two orthogonal *attB* sites (*attB1* and *attB2*) is mixed with a 'destination vector' (labelled dest.) containing a bacterial suicide gene (for example, the suicide gene *ccdB*), flanked by two related *attP* sites (*attP1* and *attP2*) that force selection of recombination between the two vectors. In the presence of the appropriate integrase enzymes, the two sites recombine to swap genes and form the hybrid sites *attR* and *attL*. Undesired assemblies (faded) are selected against by the expression of a suicide gene and by alternating antibiotic selection between survival with ampicillin resistance (*ampR*) and survival with kanamycin resistance (*kanR*). **b** | A diagram of the HomeRun Vector Assembly System (HVAS)[22] is shown. The first tier of assembly uses a Gateway reaction to build genes from elementary parts stored in Gateway entry vectors, which carry a specific pair of orthogonal recombination sites assigned to each part type (that is, promoters, open reading frames (ORFs) and terminators). Genes are assembled in one of four special Gateway destination vectors (orange) containing matching recombination sites, where they are flanked by two specific homing endonuclease (HE) sites (from a selection of four HE possibilities (HE1–HE4)). In the second tier of assembly (HomeRun cloning), genes are excised from these sites by HE digestion and sequentially ligated into a final destination vector (yellow) containing all four HE sites.

especially if repetitive DNA parts are present. Such requirements define the many standards that have emerged for these methods. The most straightforward standards have been developed for specific tasks. For example, DNA assembler performs the assembly of metabolic pathways in *S. cerevisiae* by initially building individual genes from elementary parts using OE-PCR. These genes are built with homologous ends so that, upon their transformation into yeast cells, they recombine to form the desired metabolic pathway cluster[53]. By contrast, the MODAL standard (FIG. 3b) for modular construction is more flexible and works with methods as diverse as *S. cerevisiae in vivo* assembly, Gibson Assembly and CPEC[9]. In MODAL, parts are first standardized into modules by flanking each part with universal prefix and suffix sequences. Next, the parts can be amplified using PCR with a set of orthogonal primers to add designed homology region 'linkers' that define the position and orientation of any part in the final construct, which can then be assembled by any of the techniques mentioned above (FIG. 3b). The linkers are computationally designed[54] to ensure that they are orthogonal and comply with homology requirements that ensure efficient assembly. An alternative approach avoids the use of PCR; instead, the linkers are added by cloning the parts into a plasmid in which the multiple cloning site is already flanked by the appropriate linker sequences. The linker-flanked parts can be excised using restriction endonucleases that cut outside the linkers, and the linear DNA can then be constructed using Gibson Assembly[55].

Although linkers provide many benefits, they also end up as foreign sequences between parts. This can be overcome by instead using 'bridges', which are oligonucleotides that overlap with the ends of the two parts that are going to be joined, in assembly reactions (FIG. 3c). PaperClip is a standard that can work with Gibson Assembly using a bridge-based approach[56] in which bridges with a short 3 bp scar are created by linking together short oligonucleotides that have homology with the end of each part. It is also the only standard that is currently compatible with ligase cycling reaction (LCR) assembly, a unique overlap method that uses bridges for its mechanism (FIG. 3c). In LCR, parts that are going to be joined are mixed with bridge oligonucleotides in the presence of a thermostable DNA ligase and subjected to denaturing and annealing cycles. This brings consecutive parts together and fuses them at high temperatures to ensure base-pair specificity. As fused parts act as the template for the annealing and ligation of more parts, the reaction is efficient, and a successful one-step assembly of as many as 20 parts has been reported[57].

Recent developments are now beginning to combine principles from different assembly strategies. Biopart assembly standard for idempotent cloning (BASIC) is a development of MODAL, in that orthogonal linkers are used to guide overlap assembly, but type IIS restriction endonuclease digestion is used instead of PCR to attach these to DNA parts[58]. This allows the efficiency of long-overlap methods to be combined with the simplicity of type IIS cloning and maintains a simple single-tier structure without intermediate cloning steps. Furthermore, methylation can be used to protect linkers from digestion, enabling an idempotent format for hierarchical construction through further rounds of assembly.

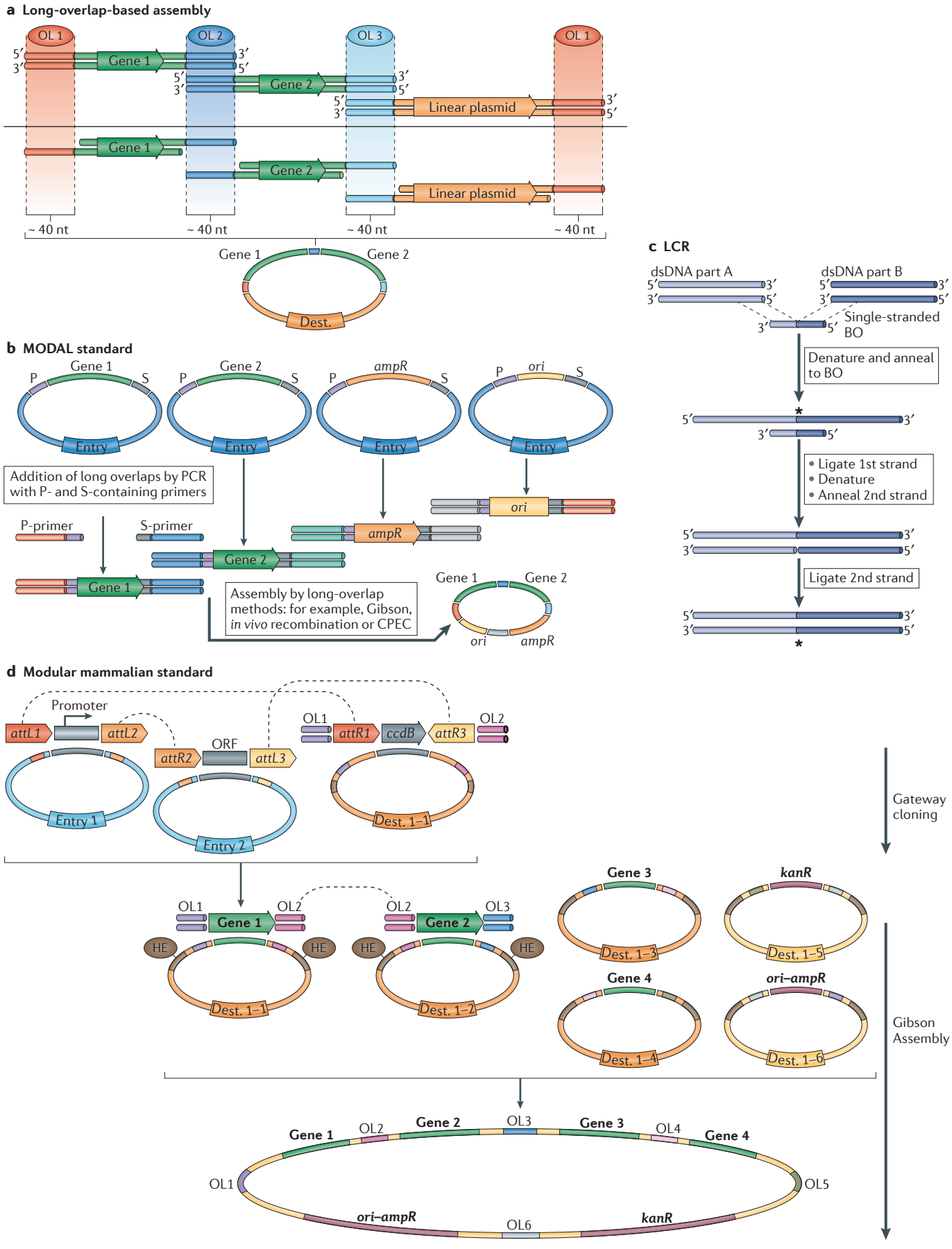
Integration of multiple methods is also key to the modular mammalian standard for the construction of large multigene plasmids for mammalian cell experiments[59]. This begins with an approach similar to the HVAS standard in which Gateway is used to assemble elementary parts into genes on plasmids, which in this case have designed linker sequences flanking the integration sites. Assembled genes flanked by overlapping linkers are then excised by homing endonuclease digestion and further assembled using Gibson Assembly to place multiple genes into a destination plasmid that can be transfected into cells (FIG. 3d). The assembled constructs maintain their linker sequences and homing endonuclease sites, which upon excision leave behind large scars but also enable further rounds of assembly without any need for PCR or consideration of forbidden restriction sites. This standard has already permitted the assembly of constructs >60 kb with more than 30 parts, enabling researchers to quickly build novel genetic programmes composed of more than 10 interacting genes and to test them in mammalian cells[59].

Conclusion and future perspectives

The variety of the new assembly methods and standards described here is already being put to use globally for impressively diverse tasks. Golden Gate assembly is at the centre of many of the kits for genome editing (BOX 1), so numerous laboratories will already be familiar with its ability to assemble many different parts in a defined order in a single reaction. Similarly, Gibson Assembly has become the method of choice in synthetic biology[60] and has even been used to assemble an entire mouse mitochondrial genome from hundreds of oligonucleotides[61]. Implementing standards for large DNA assembly projects is also beginning to yield results. By using a variation of the MoClo Golden Gate standard, researchers were recently able to automate the design and construction of 122 different versions of a cluster of 16 genes for nitrogen fixation, building from a starting library of 103 parts[62]. At the largest scale of DNA assembly, the landmark genome synthesis projects for *Mycoplasma genitalium*[63] and *Mycoplasma mycoides*[64] have shown that different scales of assembly require different methods: Gibson Assembly can be used to join DNA fragments the size of genes on a scale of ≤100 kb, but *in vivo* recombination assembly in *S. cerevisiae* becomes necessary when working on a larger scale. The global project to construct a synthetic version of the yeast genome also highlights the need for different methods at different scales and utilizes combinations of Gibson Assembly, USER cloning, traditional digestion and ligation reactions and *in vivo* recombination to hierarchically combine short DNA fragments into 50 kb synthetic 'megachunks' that replace their equivalent endogenous regions in the genome[65,66]. Given that the work of this project is shared between teams around the world, it is not surprising that standardization is required to ensure efficient progress.

Recognizing that different methods work best at different scales is already leading to a new generation of assembly approaches in which standard workflows integrate multiple methods together. By formatting parts and protocols according to a standard, future workflows will be set so that parts can be combined efficiently over sequential tiers and exchanged between projects around the world. As these standards are implemented, the next logical steps for DNA assembly methods should recapitulate those for established engineering disciplines: the development of enabling software, the introduction of automated labour and the ultimate commercial outsourcing of cloning work — as has already occurred for gene synthesis and DNA sequencing methods.

Software tools will have an increasingly important role in DNA assembly and are already required for the design of parts for modular DNA assembly[54,55,59] and used in the experimental planning and quality control of large and complex projects[67–69]. Many of these software tools can also control



◂ Figure 3 | **Examples of long-overlap-based assembly methods and standards.** **a** | A schematic of long-overlap assembly is shown. Parts for assembly have flanking overlap (OL) regions that are typically 40 nucleotides in length and define the final arrangement of the construct. Parts such as genes and the linearized vector are joined by either *in vitro* enzymatic reactions or *in vivo* recombination. The schematic depicts *in vitro* methods such as Gibson Assembly[2] (top) or sequence- and ligation-independent cloning (SLIC)[48] (bottom), in which exonuclease digestion creates single-stranded overhang overlaps that anneal to each other. **b** | In the modular overlap-directed assembly with linkers (MODAL) standard[9], both genes and backbone parts such as the ampicillin resistance gene (*ampR*) and an origin of replication (*ori*) sequences are held in entry vectors and flanked by prefix (P) and suffix (S) sequences. These are amplified by PCR with standard primers fused to 'linkers' so that the parts flanked by linkers can be assembled into a vector by various long-overlap assembly methods: for example circular polymerase extension cloning (CPEC). **c** | A schematic of ligase cycling reaction (LCR) assembly using a 'bridge'-based approach is shown[57]. This method uses single-stranded bridge oligonucleotides (BOs) complementary to the ends of two DNA parts (A and B) that are being assembled. Temperature cycling denatures parts and anneals them to the provided BOs. A thermostable ligase then joins the DNA at the bridge (marked with an asterisk). **d** | The modular mammalian standard[59] uses Gateway cloning to build genes from elementary parts (for example, promoters and open reading frames (ORFs)) held in entry vectors, which during the process of assembly are placed into 'destination vectors' (labelled dest.) that carry OL sequences which guide the second tier of assembly. Genes with OL sequences are then excised from these vectors using homing endonuclease (HE) digestion and, using Gibson Assembly, are combined in a one-pot reaction into a final vector by joining the genes as well as two linearized vector fragments (yellow) that provide antibiotic selection (through the kanamycin resistance gene (*kanR*)) and bacterial antibiotic selection and vector replication (*ori–ampR*). dsDNA, double-stranded DNA.

liquid-handling robots, and their ability to automate hundreds of complex modular assemblies has been recently demonstrated[62]. Not surprisingly, companies that now sell advanced DNA assembly software and companies that perform outsourced DNA assembly have emerged. Such companies also sell laboratory automation equipment specifically designed to run new DNA assembly methods. As the cloning of everything from plasmids to whole chromosomes by DNA assembly methods moves towards becoming a commercial service, it is likely that further advances will specifically focus on decreasing the cost while increasing the efficiency and fidelity. Dramatic improvements on all of these fronts may be achieved by incorporating microfluidics into workflows or by replacing enzymatic steps with physical and chemical routes for DNA assembly. However, it is more likely that the key to success for future DNA assembly methods will be the efficient exploitation of cheap gene and oligonucleotide synthesis combined with smart use of low-cost next-generation sequencing.

Arturo Casini, Marko Storch and Geoffrey S. Baldwin are at the Centre for Synthetic Biology and Innovation, Department of Life Sciences, Imperial College London, London SW7 2AZ, UK.

Tom Ellis is at the Centre for Synthetic Biology and Innovation, Department of Bioengineering, Imperial College London, London SW7 2AZ, UK.

Correspondence to T.E. and G.S.B
e-mails: t.ellis@imperial.ac.uk; g.baldwin@imperial.ac.uk

doi:10.1038/nrm4014
Published online 17 June 2015

1. Kosuri, S. & Church, G. M. Large-scale *de novo* DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507 (2014).
2. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
3. Keasling, J. D. Synthetic biology for synthetic chemistry. *ACS Chem. Biol.* **3**, 64–76 (2008).
4. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12**, 381–390 (2014).
5. Czar, M. J., Anderson, J. C., Bader, J. S. & Peccoud, J. Gene synthesis demystified. *Trends Biotechnol.* **27**, 63–72 (2009).
6. Annaluru, N. *et al.* Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55–58 (2014).
7. Ellis, T., Adie, T. & Baldwin, G. S. DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol.* **3**, 109–118 (2011).
8. Chao, R., Yuan, Y. & Zhao, H. Recent advances in DNA assembly technologies. *FEMS Yeast Res.* **15**, 1–9 (2015).
9. Casini, A. *et al.* One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. *Nucleic Acids Res.* **42**, e7 (2014).
10. Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology — identifying causes of failure of synthetic biological systems. *Biotechnol. J.* **7**, 856–866 (2012).
11. Cohen, S. N., Chang, A. C., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl Acad. Sci. USA* **70**, 3240–3244 (1973).
12. Crook, N. C., Freeman, E. S. & Alper, H. S. Re-engineering multicloning sites for function and convenience. *Nucleic Acids Res.* **39**, e92 (2011).
13. Shetty, R. P., Endy, D. & Knight, T. F. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2**, 5 (2008).
14. Norville, J. E. *et al.* Introduction of customized inserts for streamlined assembly and optimization of BioBrick synthetic genetic circuits. *J. Biol. Eng.* **4**, 17 (2010).
15. Anderson, J. C. *et al.* BglBricks: a flexible standard for biological part assembly. *J. Biol. Eng.* **4**, 1 (2010).
16. Leguia, M., Brophy, J. A., Densmore, D., Asante, A. & Anderson, J. C. 2ab assembly: a methodology for automatable, high-throughput assembly of standard biological parts. *J. Biol. Eng.* **7**, 2 (2013).
17. Litcofsky, K. D., Afeyan, R. B., Krom, R. J., Khalil, A. S. & Collins, J. J. Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. *Nat. Methods* **9**, 1077–1080 (2012).
18. Silva-Rocha, R. *et al.* The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* **41**, D666–D675 (2013).
19. Martínez-García, E., Aparicio, T., Goñi-Moreno, A., Fraile, S. & de Lorenzo, V. SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res.* **43**, D1183–D1189 (2014).
20. Marcaida, M. J., Muñoz, I. G., Blanco, F. J., Prieto, J. & Montoya, G. Homing endonucleases: from basics to therapeutic applications. *Cell. Mol. Life Sci.* **67**, 727–748 (2010).
21. Liu, J.-K., Chen, W.-H., Ren, S.-X., Zhao, G.-P. & Wang, J. iBrick: a new standard for iterative assembly of biological parts with homing endonucleases. *PLoS ONE* **9**, e110852 (2014).
22. Li, M. V. *et al.* HomeRun vector assembly system: a flexible and standardized cloning system for assembly of multi-modular DNA constructs. *PLoS ONE* **9**, e100948 (2014).
23. Sleight, S. C. & Sauro, H. M. Visualization of evolutionary stability dynamics and competitive fitness of *Escherichia coli* engineered with randomized multi-gene circuits. *ACS Synth. Biol.* **2**, 519–528 (2013).
24. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3**, e3647 (2008).
25. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39**, e82 (2011).
26. Kamens, J. The Addgene repository: an international nonprofit plasmid and data resource. *Nucleic Acids Res.* **43**, D1152–D1157 (2015).
27. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10**, 354–360 (2013).
28. Linshiz, G. *et al.* PR-PR: cross-platform laboratory automation system. *ACS Synth. Biol.* **3**, 515–524 (2014).
29. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* **6**, e16765 (2011).
30. Sarrion-Perdigones, A. *et al.* GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol.* **162**, 1618–1631 (2013).
31. Binder, A. *et al.* A modular plasmid assembly kit for multigene expression, gene silencing and silencing rescue in plants. *PLoS ONE* **9**, e88218 (2014).
32. Lampropoulos, A. *et al.* GreenGate — a novel, versatile, and efficient cloning system for plant transgenesis. *PLoS ONE* **8**, e83043 (2013).
33. Sarrion-Perdigones, A. *et al.* GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS ONE* **6**, e21622 (2011).
34. Duportet, X. *et al.* A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res.* **42**, 13440–13451 (2014).
35. Sakuma, T., Nishikawa, A., Kume, S., Chayama, K. & Yamamoto, T. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci. Rep.* **4**, 5400 (2014).
36. Chen, W.-H., Qin, Z.-J., Wang, J. & Zhao, G.-P. The MASTER (methylation-assisted tailorable ends rational) ligation method for seamless DNA assembly. *Nucleic Acids Res.* **41**, e93 (2013).
37. Blake, W. J. *et al.* Pairwise selection assembly for sequence-independent construction of long-length DNA. *Nucleic Acids Res.* **38**, 2594–2602 (2010).
38. Hartley, J. L., Temple, G. F. & Brasch, M. A. DNA cloning using *in vitro* site-specific recombination. *Genome Res.* **10**, 1788–1795 (2000).
39. Alberti, S., Gitler, A. D. & Lindquist, S. A suite of Gateway cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast* **24**, 913–919 (2007).
40. Marsischky, G. & LaBaer, J. Many paths to many clones: a comparative look at high-throughput cloning methods. *Genome Res.* **14**, 2020–2028 (2004).
41. Sasaki, Y. *et al.* Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the multisite Gateway system. *J. Biotechnol.* **107**, 233–243 (2004).
42. Zhang, L., Zhao, G. & Ding, X. Tandem assembly of the epothilone biosynthetic gene cluster by *in vitro* site-specific recombination. *Sci. Rep.* **1**, 141 (2011).

43. Colloms, S. D. *et al.* Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res.* **42**, e23 (2013).
44. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**, 61–68 (1989).
45. Quan, J. & Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE* **4**, e6441 (2009).
46. Bitinaite, J. *et al.* USER friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Res.* **35**, 1992–2002 (2007).
47. Annaluru, N. *et al.* Assembling DNA fragments by USER fusion. *Methods Mol. Biol.* **852**, 77–95 (2012).
48. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256 (2007).
49. Zhang, Y., Werling, U. & Edelmann, W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* **40**, e55 (2012).
50. Itaya, M., Fujita, K., Kuroki, A. & Tsuge, K. Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nat. Methods* **5**, 41–43 (2008).
51. Raymond, C. K., Pownder, T. A. & Sexson, S. L. General method for plasmid construction using homologous recombination. *Biotechniques* **26**, 134–141 (1999).
52. Gibson, D. G. *et al.* One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl Acad. Sci. USA* **105**, 20404–20409 (2008).
53. Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* **37**, e16 (2009).
54. Casini, A. *et al.* R2oDNA designer: computational design of biologically neutral synthetic DNA sequences. *ACS Synth. Biol.* **3**, 525–528 (2014).
55. Torella, J. P. *et al.* Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly. *Nucleic Acids Res.* **42**, 681–689 (2013).
56. Trubitsyna, M., Michlewski, G., Cai, Y., Elfick, A. & French, C. E. PaperClip: rapid multi-part DNA assembly from existing libraries. *Nucleic Acids Res.* **42**, e154 (2014).
57. de Kok, S. *et al.* Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synth. Biol.* **3**, 97–106 (2014).
58. Storch, M. *et al.* BASIC: a new biopart assembly standard provides accurate, single-tier DNA assembly for synthetic biology. *ACS Synth. Biol.* http://dx.doi.org/10.1021/sb500356d (2015).
59. Guye, P., Li, Y., Wroblewska, L., Duportet, X. & Weiss, R. Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res.* **41**, e156 (2013).
60. Kahl, L. J. & Endy, D. A survey of enabling technologies in synthetic biology. *J. Biol. Eng.* **7**, 13 (2013).
61. Gibson, D. G., Smith, H. O., Hutchison, C. A., Venter, J. C. & Merryman, C. Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods* **7**, 901–903 (2010).
62. Smanski, M. J. *et al.* Functional optimization of gene clusters by combinatorial design and assembly. *Nat. Biotechnol.* **32**, 1241–1249 (2014).
63. Gibson, D. G. *et al.* Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
64. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
65. Jovicevic, D., Blount, B. A. & Ellis, T. Total synthesis of a eukaryotic chromosome: redesigning and SCRaMbLE-ing yeast. *Bioessays* **36**, 855–860 (2014).
66. Dymond, J. S. *et al.* Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011).
67. Hillson, N. J., Rosengarten, R. D. & Keasling, J. D. j5 DNA assembly design automation software. *ACS Synth. Biol.* **1**, 14–21 (2012).
68. Appleton, E., Tao, J., Haddock, T. & Densmore, D. Interactive assembly algorithms for molecular cloning. *Nat. Methods* **11**, 657–662 (2014).
69. Densmore, D. *et al.* Algorithms for automated DNA assembly. *Nucleic Acids Res.* **38**, 2607–2616 (2010).
70. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
71. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).
72. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
73. Pósfai, G. *et al.* Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**, 1044–1046 (2006).
74. Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15**, 321–334 (2014).
75. Horwitz, A. A. *et al.* Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR–Cas. *Cell Syst.* http://dx.doi.org/10.1016/j.cels.2015.02.001 (2015).
76. Groth, A. C. & Calos, M. P. Phage integrases: biology and applications. *J. Mol. Biol.* **335**, 667–678 (2004).
77. Torella, J. P. *et al.* Unique nucleotide sequence-guided assembly of repetitive DNA parts for synthetic biology applications. *Nat. Protoc.* **9**, 2075–2089 (2014).

Acknowledgements
The authors thank M. Jamilly for contributing to initial discussions and apologize to those whose work was not discussed owing to space limitations. Research on DNA assembly in the groups of T.E. and G.S.B. is supported by the UK Engineering and Physical Research Council (EPSRC) grant EP/J02175X/1 and EU FP7 grant KBBE.2011.5-289326. M.S. is supported by Marie Curie Intra-European Fellowship 628019.

Competing interests statement
The authors declare no competing interests.