ABSTRACT Since the discovery of restriction enzymes and the generation of the first recombinant DNA molecule over 40 years ago, molecular biology has evolved into a multidisciplinary field that has democratized the conversion of a digitized DNA sequence stored in a computer into its biological counterpart, usually as a plasmid, stored in a living cell. In this article, we summarize the most relevant tools that allow the swift assembly of DNA sequences into useful plasmids for biotechnological purposes. We cover the main components and stages in a typical DNA assembly workflow, namely in silico design, de novo gene synthesis, and in vitro and in vivo sequence assembly methodologies.

INTRODUCTION
Recombinant plasmids are possibly the biological reagents most frequently used in molecular biology. Myriad plasmid assembly strategies have been devised since the first recombinant DNA molecule was generated over 40 years ago (1). Cloning protocols are now so profuse that it is not always trivial to choose one that is the most suitable for a particular purpose. The aspects to consider at the time of choosing an assembly strategy include, among others, the nature of the sequences, fragment size and number, template availability, plasmid capacity and stability, and selection against a background of other unwanted assemblies.

Before the emergence of PCR in the late 1980s, recombinant DNA approaches were based exclusively on restriction enzyme-based protocols. These sequence-dependent procedures rely on unique and specific sites in the target molecules and do not allow the seamless cloning of various DNA fragments at the same time. During the past two decades, numerous sequence-dependent and -independent strategies have been developed that permit the simultaneous assembly of multiple sequences of various sizes with high efficiencies, flexible throughputs, and reduced hands-on and overall times.

In this article, we attempt to provide a summary of useful tools for constructing plasmids, emphasizing on those that are most recent and innovating, stating their advantages and disadvantages. We refer the reader to complementary articles for those technologies not mentioned or reviewed here in sufficient detail (2, 3, 4, 5, 6, 7).

CHOICE OF REPLICON
Usually, three major factors are the determinants of what type of plasmid to use: (i) the downstream application, (ii) the size of the expected construct, and
Baek et al.

(iii) the desired copy number. If the cloning approach (see below) is predetermined, then it would also play a significant role in the decision. The plasmid type is largely imposed by the nature of its origin of replication (replicon). In this section, we intend to provide examples of available replicon alternatives geared toward the categories below. For deeper information on origins of replication, we suggest consulting earlier literature (8, 9, 10).

By far, the most common replicon in *Escherichia coli* is the ColE1 type derived from the plasmid pMB1, which evolved via plasmids pBR322 and pUC18/19 (11, 12). Thus, vectors containing ColE1/pMB1-derived origins of replication fall into the same incompatibility group (13). The major advantages of these replicons are their multicopy nature, which ensures high plasmid DNA yield and makes them desirable for many downstream processes. On the other hand, derivatives of these plasmids have a limited host range and capacity, usually between 15 and 20 kb, which may hinder their value in applications that require large genetic elements as is the case of the construction of circuits for synthetic biology. Other shortcomings are the potential side effects of their multicopy nature, such as (i) the toxicity of certain plasmid elements, or (ii) the difficulty on the downregulation of gene expression, necessary to balance the expression of different components in complex gene networks.

These shortcomings could be avoided by using compatible replicons with a different copy number, host range, and capacity (Table 1). Comprehensive reviews on functional replicons for *E. coli* and other Gram-negative bacteria have been published previously (8, 14). When copy number must be strictly limited to 1, the solutions are either (i) to recombine the DNA fragment into the bacterial chromosome, or (ii) to use the *E. coli* F factor (Table 1) (for a review see reference 10). The latter option has the advantage of being a large-size-capacity conjugative episome. However, the host range is limited to *E. coli* and some related bacteria. For broad-range Gram-negative host options, plasmids from the IncP and IncQ incompatibility groups are recommended, yet their size limits are in the order of a few ten thousand base pairs (Table 1).

The emerging synthetic biology and bioengineering fields require the cloning of large constructs, such as genetic pathways or even chromosomes that, because of their sheer size and number of input segments, are difficult (often impossible) to assemble in *E. coli*. In these cases, the only viable cloning organism has been shown to be *Saccharomyces cerevisiae* (Table 1). Large sequences can be cloned circularly in yeast as circular episomes or yeast artificial chromosomes (YACs). The two major functional requirements of these elements are (i) a yeast centromere (CEN) and (ii) an autonomously replicating sequence (ARS) (for a review, see reference 15). By combining different CEN and ARS elements, stable yeast episomes can be used to clone sequences in the mega base pair scale (Table 1) (16).

**SEQUENCE-DEPENDENT CLONING IN VITRO**

A common feature of sequence-dependent cloning approaches is the requirement of specific sites in the inserts and/or vector. Many of these leave behind operational sequences that could interfere with the function of the genetic assembly.

The classical example of a sequence-dependent cloning strategy is that of one built upon restriction enzymes

---

**TABLE 1** Replicons most commonly used in plasmids

<table>
<thead>
<tr>
<th>Replicons</th>
<th>Related vectors</th>
<th>Copy number</th>
<th>Capacity</th>
<th>Host</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMB1/ColE1</td>
<td>pBR322 and derivatives</td>
<td>15–20</td>
<td>Up to, at least 15 kb</td>
<td><em>E. coli</em></td>
<td>11</td>
</tr>
<tr>
<td>Mutated pMB1</td>
<td>pUC vectors</td>
<td>500–700</td>
<td>Up to, at least 15 kb</td>
<td><em>E. coli</em></td>
<td>12</td>
</tr>
<tr>
<td>p15A</td>
<td>pACYC184</td>
<td>18–22</td>
<td>Up to, at least 10 kb</td>
<td><em>E. coli</em></td>
<td>80</td>
</tr>
<tr>
<td>pSC101</td>
<td>pSC101</td>
<td>Approximate 5</td>
<td>Up to, at least 16 kb</td>
<td><em>E. coli</em></td>
<td>81</td>
</tr>
<tr>
<td>IncP</td>
<td>RK2</td>
<td>4–40 (adjustable by mutations)</td>
<td>Up to, at least 60 kbp</td>
<td>Broad host range</td>
<td>82, 83</td>
</tr>
<tr>
<td>IncQ</td>
<td>RSF1010</td>
<td>12</td>
<td>Up to, at least 14 kbp</td>
<td>Broad host range</td>
<td>84</td>
</tr>
<tr>
<td>R6K</td>
<td>R6K and its derivatives</td>
<td>15–30</td>
<td>Up to, at least 38 kbp</td>
<td><em>E. coli</em></td>
<td>85</td>
</tr>
<tr>
<td>F factor</td>
<td>F plasmid and its derivatives (i.e., bacterial artificial chromosomes [BACs])</td>
<td>1</td>
<td>Often up to 350 kb</td>
<td><em>E. coli</em></td>
<td>30</td>
</tr>
<tr>
<td>2-micron</td>
<td>2-micron and derivatives</td>
<td>Approximately 100</td>
<td>At least 3.9 kb</td>
<td><em>S. cerevisiae</em></td>
<td>86</td>
</tr>
<tr>
<td>Cen6/ARS4</td>
<td>pRS series</td>
<td>Possibly 1</td>
<td>1–2 Mbp</td>
<td><em>S. cerevisiae</em></td>
<td>51, 87</td>
</tr>
<tr>
<td>Cen4/ARS1</td>
<td>pYAC2, pYAC4</td>
<td>&gt;12</td>
<td>&gt;1 Mbp</td>
<td><em>S. cerevisiae</em></td>
<td>88, 89</td>
</tr>
</tbody>
</table>
and ligases. Although this method is allegedly the most widely used tool in any molecular biology laboratory, it is relatively inefficient in the sense that (i) it is based on particular target sequences, (ii) it requires multiple stages and verification steps, (iii) it is not universal because each particular experiment uses a specific enzyme and conditions, (iv) it leaves behind scar sequences, and (v) it can assemble a very limited number of fragments at once. Given the vast array of information accumulated on this topic, we recommend consulting earlier publications (5, 7, 17).

With the emergence of PCR, a number of sequence-dependent methods were proposed, such as TA cloning (18), topoisomerase-based cloning (TOPO cloning) (19), and site-specific recombination approaches such as Gateway (20, 21) and Cre/LoxP recombination (22). Main providers for commercial products using the methodologies above are New England Biolabs (Ipswich, MA), Agilent Technologies (Santa Clara, CA), and Life Technologies (Carlsbad, CA). For further details on these technologies, refer to earlier reviews (23, 24).

More recently, restriction ligation cloning has found an application that standardizes the DNA assembly process, the BioBrick assembly (25). Each BioBrick part consist of a DNA sequence flanked by two different and unique restriction sites, which allows the generation of larger BioBrick parts by chaining together smaller ones. This process generates scar junctions that lack the original sites, still allowing for idempotent cloning cycles. A bundled BioBrick Assembly Kit is being commercialized by New England Biolabs (Ipswich, MA). In some specific cases, the creation of seamless connections between arbitrary DNA sequences in cloning vectors can be established by using a BioBrick part (called a BioScaffold) that can be excised by a type IIB restriction enzyme to leave a gap into which other DNA elements can be cloned (26).

Another way to resolve the presence of undesired scars in the final construct is by the use of type IIS restriction enzymes. These endonucleases cleave the DNA at distal locations from the recognition sites. The recognition sites can be strategically placed into regions of PCR-amplified fragments that will be excluded from the final construct, and, therefore, the restriction and ligation steps can be consolidated into a single reaction (27, 28). The strategy requires that these particular target sites must be absent in the individual parts. The method has been termed “Golden Gate cloning” (Fig. 1A) and has been used for (i) assembling gene variants or repetitive sequences (29, 30), (ii) generating iterative cloning systems (31), and (iii) obtaining standardized parts for multigene assemblies (32). Kits using this technology accompanied by a design web tool are being commercialized under the name of GeneArt type IIs by Life Technologies (Carlsbad, CA). Variants of this method have been proposed where the type IIS endonuclease is replaced either by a nicking enzyme (33) or by a type IIM restriction enzyme (34). In this latter case, the enzyme recognizes a methylated site, virtually transforming this approach into a sequence-independent strategy.

### SEQUENCE-INDEPENDENT CLONING IN VITRO

During the past few years, a number of methods that do not rely on specific target sites and do not leave unwanted sequences in the final construct have been developed. Some of them are based on the ability of exonucleases to “chew back” one of the strands of double-stranded (ds) DNA, thereby exposing complementary single-stranded (ss) DNA sequences that can anneal to each other (Fig. 1B). The remaining gaps or nicks can be repaired in vitro or within the host organism. Examples of these include ligation-independent cloning (LIC) (35), sequence-and-ligation-independent cloning (SLIC) (36), fusion assembly (37), isothermal/Gibson assembly (38), and GeneArt Seamless and Cloning Assembly (39). Several of these technologies have been turned into products, commercialized by a variety of companies such as New England Biolabs (Ipswich, MA), EMD (San Diego, CA), Clontech (Mountain View, CA), and Life Technologies (Carlsbad, CA).

A similar recombination principle is used by systems that rely on different modes of exposing single-strand complementary sequences. For example, the uracil-specific excision reagent (USER), composed of uracil DNA glycosylase and endonuclease VIII, catalyzes the excision of a deoxyuracil residue, incorporated into a PCR fragment by a corresponding oligonucleotide (40) (Fig. 1C). In a different approach, DNA amplification is performed by using oligonucleotides containing phosphorothioate nucleotides in which a phosphodiester bond is replaced by a phosphothioester bond. These bonds can be chemically cleaved in an iodine/ethanol solution at elevated temperatures (41, 42). By strategically positioning the modified nucleotides in both methods above, fragments with compatible recessed ends can be generated and self-organized in a predefined order. Finally, in a related approach, compatible ssDNA
overhangs are generated by PCR amplification with the use of oligonucleotides that contain regions of RNA sequence that cannot be copied by certain thermostable DNA polymerases (43). From the three approaches above, only the USER method is commercially available (marketed by New England Biolabs, Ipswich, MA).

A third approach relies entirely on polymerase extension of one or multiple overlapping fragments without calling for ssDNA complementary regions for recombination. In its simplest conception, a DNA fragment can be seamlessly inserted into a recipient vector by two consecutive PCR amplification assays (44). During the first one, tails complementary to the insertion region are added to the fragment to be inserted. The resulting dsDNA fragment is used as a giant primer to PCR amplify the recipient plasmid, thereby generating the desired molecule. This method can be used for site-directed mutagenesis as well (see below). In a more recent related idea, it was shown that a similar polymerase extension strategy can be applied to join overlapping

FIGURE 1 Schematic representation of different DNA assembly methodologies. (A) Golden Gate Cloning. Fragments to be assembled (red and green) have strategically placed terminal type IIS endonuclease recognition sites (in this case BsaI sites shown in lowercase and underlined). Black arrowheads point toward the BsaI cleavage sites. Simultaneous incubation with BsaI and DNA ligase results in covalently linked fragments. (B) Chew-back and repair-based assembly. Adjacent DNA fragments (red and green) sharing terminal sequence overlaps are incubated with DNA exonuclease, thereby exposing complementary DNA strands. The strands are annealed and the gaps can be sealed either in vitro by DNA polymerase and DNA ligase, or by the cell upon transformation. (C) USER assembly. Adjacent fragments (red and green), amplified with compatible uracil-containing primers, are incubated with the USER enzyme mix, which removes the uracils. The small terminal complementary DNA strands (in black) anneal to each other, outcompeting the small terminal loose strand. Gaps are repaired and sealed by the cell upon transformation. doi:10.1128/microbiolspec.PLAS-0014-2013.f1
DNA fragments into a double-stranded circular form without a template (45).

**IN VIVO CLONING**

The “chew-back and repair” mechanism has also been applied to strategies where the assembly process occurs entirely within a living cell. Several hosts are particularly proficient in homologous recombination, a trait that could be used for cloning purposes. *E. coli* cells exhibit poor homologous recombination activity. However, the expression of the phage lambda redET genes (lambda red system) strongly promotes homologous recombination between a linear and circular DNA molecule (46) or between two linear DNA molecules (47). Gene replacement and simple cloning reactions can be performed by using *E. coli* strains harboring the redET genes, however, multiple fragments cannot easily be assembled into a single construct (unpublished data). The system is commercially available from Gene Bridges (Heidelberg, Germany).

Another organism used for cloning purposes is *Bacillus subtilis*, where DNA fragments can be naturally transformed and “stitched” together in a stepwise manner, generating large epipodes or DNA structures integrated into a plasmid or the cell’s chromosome (48). The mechanism of DNA uptake by *B. subtilis* has not completely worked out, but some reports indicate that, during the process, the incoming DNA is cleaved and generates smaller ssDNA fragments (for a review, see reference 49). These features might reduce the overall efficiency of the system compared with other in vivo cloning methods.

Possibly the most widely used in vivo cloning approach is the one involving the budding yeast *S. cerevisiae*. This organism has the ability to assemble and maintain constructs larger than 1 Mbp starting from dozens of overlapping DNA fragments of varying sizes with as few as 30 bp in common at their ends (50; reference 51 and references therein). The technology was first described in the early 1990s (52), but it really took off during the 2000s when further applications were reported. For example, it was shown that the system can join DNA fragments with non-homologous ends by means of double-stranded oligonucleotides (53). This property allows the reuse of existing fragments without the need to reamplify them to incorporate end-homology. More recently, it has been demonstrated that yeast can also assemble constructs starting from overlapping oligonucleotides, with overlaps as short as 20 bp (54). Depending on the final size, assembled constructs can be transferred back to *E. coli* by “electroporating” lysed yeast colonies into electrocompetent *E. coli* cells (39). Finally, the combination of homing endonuclease recognition sites with compatible yeast markers has been shown to allow the sequential insertion of an indefinite number of DNA fragments in a predetermined locus (55). The *S. cerevisiae* recombination system is commercialized by Life Technologies (Carlsbad, CA) under the name of GeneArt High-Order Genetic Assembly System.

**SITE-DIRECTED MUTAGENESIS**

In some plasmid construction workflows, rather than reassembling the episome from a variety of constituent parts, it is much simpler to apply minor modifications to an existing molecule. During the past three decades, site-directed mutagenesis has become one of the most powerful tools in genetics. Its power lies in its ability, by chemical and/or enzymatic manipulation, to change a specific DNA target in a definable and often predetermined way. With the advent of synthetic biology and rational design, the manipulation of genes to produce enzymes with subtle differences from wild type has increased significantly. An array of methods has been described in the literature (for extensive information on this topic, see reference 56). Several of these technologies are commercialized by a variety of companies (for some examples, see Table 2). These products use at least one of the following approaches: (i) the isolation of a single- or double-stranded circular DNA template to create the mutation by using one or more primers that anneal to the same strand (57, 58, 59, 60); (ii) the design of two sets of PCR primers that overlap the mutation site (57, 59, 60, 61); (iii) the PCR amplification of a plasmid by using complementary oligonucleotides and the subsequent elimination of the template molecule (44, 62, 63); (iv) a ligation during amplification strategy by inward PCR amplification of circular templates using phosphorylated primers (64); (v) the divergent PCR amplification of circular templates by using phosphorylated primers followed by ligation; and (vi) the *in vitro* recombination of the ends of one or more PCR fragments (65). Preferred systems are those that use unmodified desalted oligonucleotides, dsDNA as a substrate, and a minimum number of steps. Some of the products listed in Table 2 work for single-site and multi-site mutagenesis. For further details, we suggest consulting the particular manufacturer’s protocols.
TABLE 2 Examples of commercial site-directed mutagenesis kits

<table>
<thead>
<tr>
<th>Name of the product</th>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change-IT Multiple Site Directed Mutagenesis Kit</td>
<td>Affymetrix (Santa Clara, CA)</td>
<td>64</td>
</tr>
<tr>
<td>QuikChange Lightning Site Directed Mutagenesis Kit</td>
<td>Agilent Technologies (Santa Clara, CA)</td>
<td>59, 60, 62, 63</td>
</tr>
<tr>
<td>Phusion Site-Directed Mutagenesis Kit</td>
<td>Thermo Fisher (Waltham, MA) and New England Biolabs (Ipswich, MA)</td>
<td></td>
</tr>
<tr>
<td>GeneArt Site-Directed Mutagenesis Kits</td>
<td>Life Technologies (Carlsbad, CA)</td>
<td>65</td>
</tr>
<tr>
<td>Altered Sites II in vitro Mutagenesis System</td>
<td>Promega (Madison, WI)</td>
<td>57, 58</td>
</tr>
<tr>
<td>GeneEditor in vitro Site-Directed Mutagenesis System</td>
<td>Promega (Madison, WI)</td>
<td>57, 58</td>
</tr>
</tbody>
</table>

DE NOVO GENE SYNTHESIS

What options, other than those mentioned above, are available for scenarios where (i) no DNA templates are available or (ii) the most straightforward cloning strategy to reach the desired goal is too tedious or convoluted? Perhaps the simple answer is artificial gene synthesis.

The first example of de novo synthesis of a DNA sequence was demonstrated by Khorana and coworkers in 1970 (66). In an effort that took several years, they assembled a 77-bp gene encoding a yeast alanine transfer RNA by using short oligonucleotides obtained through organic chemistry methods. In 1990, Mandecki and colleagues crossed the 1000-bp size barrier with the synthesis of a 2.1-kbp fully synthetic plasmid (67). Yet, another 20 years later, Venter’s group designed, synthesized, and assembled the 1.08-Mbp Mycoplasma mycoides JCVI-syn1.0 genome starting from digitized genome sequence information, resulting in the first self-replicating organism derived from a fully synthetic genome (68).

Today, de novo gene synthesis enhances the biotechnology field concerning: safety, availability, reliability, throughput, flexibility, and, last but not least, total cost. Its application not only completely changed the way in which scientists think when designing cloning strategies, but also facilitated the outsourcing of related experimental steps in order to concentrate on less trivial scientific operations. Researchers can electronically access DNA sequences through comprehensive databases, redesign constructs in silico to specifically fit given requirements, and then order online the designed genes being synthesized and have them shipped within a matter of days.

Given the flexibility to synthesize any conceivable string of nucleotides, it is reasonable to alter a natural gene sequence before synthesis to ensure its best performance in the required application or experiment—a process known as gene optimization. The most commonly used modification of protein-coding genes is codon usage adaptation. However, further variables to consider are (i) GC content, (ii) mRNA half-life and RNA secondary structures, avoiding direct and reverse repeats, (iii) restriction sites, (iv) ribosomal entry sites, (v) cryptic splice motifs, and (vi) polyadenylation signals, among others. Taken together, this results in a multiparameter optimization, requiring sophisticated algorithms and significant computational speed (for an example, see reference 69).

Most protocols for gene synthesis are based on the assembly of chemically synthesized oligonucleotides to yield longer contiguous sequences. Their template-less production is still a significant cost factor, and different strategies have been conceived to reduce these expenses. It is, however, particularly important not to compromise on the accuracy of the applied chemical synthesis process, because the large number of sequential chemical reactions on the elongated oligonucleotide, together with the inherent imperfection of each step, lead to an increasing probability of incorporating a mutation within the molecule. These are usually single-nucleotide deletions, insertions, or depurinations that occur with a frequency of 0.1 to 0.5% (0.1% = every 1,000th nucleotide has an error, or one of fifty 20-mers carries a mutation). When assembling many oligonucleotides into a longer contiguous molecule, the statistical clustering of mutations within a synthetic gene increases exponentially. The length of the gene and the error rate of the oligonucleotides both have a major effect on the final sequence accuracy. For example, a synthetic gene of 1,000 bp made from oligonucleotides with an error rate of 0.1% will have a total accuracy of (100% - 0.1%)1000 = 37%, while an error rate of 0.3% decreases the final accuracy to 5%. In the first case, one of three sequenced clones will contain the accurate sequence, whereas, in the second case, 20 clones will have to be sequenced, on average, to find a correct one.

Taking these factors into account, the maximal final length of the initial assembly constructs must be considered carefully, in order to find the best economical ratio between the likelihood of errors in the product and the number of transformants to screen. Currently, the most cost-effective size of these synthetic building blocks is between 1 and 3 kbp.

The assembly process itself is basically a multiplex primer extension reaction under controlled temperature...
cycling conditions (Fig. 2). In the first cycling round, overlapping primers anneal to each other and are filled in by polymerase to form short double strands. These can again anneal to each other in the subsequent cycle and are extended to fragments bridging four oligonucleotides. This progression continues until fragments arise containing the complete length of the intended product. Once achieved, the excess of terminal primers amplifies the full-length product exponentially (70).

Different techniques can then be applied to reduce the remaining errors in the assembly product before ligation and transformation. Denaturation and reannealing separates random mutations in complementary positions and results in dsDNA carrying mismatches, while strands with the correct sequence can form perfect duplex DNA. Mismatches in imperfect heteroduplexes then serve as entry points for digestion by different enzyme systems or affinity depletion by mismatch binding proteins (71, 72, 73).

After assembly, amplification, and error reduction, the linear gene synthesis product is ligated into a minimal cloning vector by using classical restriction endonuclease techniques or other assembly techniques (see preceding sections). After transformation of E. coli,

**FIGURE 2** Schematic overview of gene synthesis workflow. *In silico* designed sequence data are converted into a set of oligonucleotides by automated organic chemistry. These are stepwise assembled, elongated, and amplified into a full-length fragment (see box), which is then ligated into a cloning vector. After transformation, E. coli colonies are screened for error-free insert sequences and a correct colony is cultivated for plasmid isolation. After a final sequence verification of the plasmid preparation, the construct is ready to be used or to be further assembled into larger constructs. doi:10.1128/microbiolspec.PLAS-0014-2013.f2
some colonies are selected for plasmid preparation, and the accuracy of the synthesized DNA construct is verified by sequencing. Altogether, conditions for mass production are chosen to have a $>95\%$ chance of picking at least one correct fragment with a single screen, limiting the size of the initial fragment to 1 to 3 kbp. Compiling synthetic constructs exceeding 3 kbp requires the further sequence-independent and scarless assembly of cloned and sequence-verified building blocks (see preceding sections).

Type II class S restriction sites can produce sticky ends outside their recognition sequence, while the nucleotides of the adjacent cohesive stretch can be chosen freely, representing a common part of 4 bp for ligation (74). Designing this common part to have a length of $\sim 20$ bp allows flexible and specific attachment of two or more DNA fragments by fusion PCR, but is limited to moderate overall size and inherits an additional source of sequence errors through PCR (75). The DISEC-TRISEC and LIC-PCR methods use the exonuclease activity of Klenow or T4 DNA polymerase to generate compatible single-stranded overhangs, which are then combined with or without ligase, respectively (35, 76) (see preceding sections for additional methodologies). In vitro recombination extends this technology by annealing the overhangs under more stringent conditions at elevated temperatures and then filling and closing gaps with a heat-stable polymerase and ligase. This already allows for the efficient assembly of molecules in the range of 100 to 200 kbp (38). To access even larger fragment sizes, recent protocols have taken advantage of the recombination efficiency of yeast, enabling the assembly of complete operons and genomes of sizes exceeding 1 Mbp (see above sections).

**IN SILICO DESIGN TOOLS**

Bioinformatics software is an important component of designing and planning assembly of plasmids. Good bioinformatics tools assist users with devising, planning, and testing their ideas in silico as well as verifying and validating their experimental results. Bioinformatics software fulfills seven main needs by providing:

1. Curation capabilities—means of storing and managing data on plasmid sequences;
2. Discovery capabilities—provision of series of tools to allow users to find, compare, and analyze their sequences for biological insights;
3. Design capabilities—provision of series of tools to design or modify existing sequences, identify needed materials such as amplimers and oligonucleotides for experimental purposes, and highlight and resolve likely experimental issues in silico;
4. Inventory capabilities—provision of tools to manage their experimental materials, especially when reusing or repurposing existing plasmids, oligonucleotides, amplimers, etc.;
5. Confirmation capabilities—provision of tools to model, validate, and verify the plasmids they have created;
6. Visualization capabilities—provision of easy-to-use interfaces that allow users to easily understand what they are working on, find the tools to analyze these sequences, and transition their projects through discovery, design, and confirmation of experimental materials;
7. Sharing capabilities—coordination of groups of users working together, most simply through the use of common file formats, more extensively through user-to-user sharing of electronic records or databases.

With the recent advent of synthetic biology, scientists are increasingly using design-based approaches for developing and managing their plasmid collections (77). To achieve this, the traditional sequence and feature-based view of plasmids will be updated to regard features as discrete parts that can be reused in plasmid design projects (78). Parts, in turn, are assembled into devices that help the plasmid interact with the cell, such as antibiotic resistance devices, origin of replication devices, or expression devices (79). Devices themselves are controlled through circuits, such as the use of one or more inducer-based elements to control the expression of genes or combinations of different origins of replication to stably transform host cell lines. Bioinformatics software will need to enable users to switch back and forth between feature- and functional-based views of the same molecules.

**Table 3** contains a list of pieces of software that have been used by the scientific community for cloning-based tasks. All of these pieces of software have user interfaces that assist the end user with managing their data, and they vary most in terms of the types of tools offered for modeling different sequence assembly strategies. Due to the prevalence of restriction enzyme-based cloning, all support these types of assemblies. The main differentiators in the collection consist of the use of more recent sequence-independent cloning methodologies, highlighted by the availability of Gateway cloning, homologous recombination, and synthetic biology...
## TABLE 3 Most commonly used bioinformatics software

<table>
<thead>
<tr>
<th>Product name</th>
<th>Platform</th>
<th>Website</th>
<th>Plasmid construction support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene inspector*</td>
<td>Mac/Windows/Unix</td>
<td><a href="http://www.textco.com">http://www.textco.com</a></td>
<td>Restriction, Golden Gate</td>
</tr>
<tr>
<td>DNA Dynamo*</td>
<td>Mac/Windows/Unix</td>
<td><a href="http://www.bluetractorsoftware.co.uk/">http://www.bluetractorsoftware.co.uk/</a></td>
<td>Restriction, Golden Gate</td>
</tr>
<tr>
<td>Benchling*</td>
<td>Mac/Windows/Unix</td>
<td><a href="https://benchling.com/">https://benchling.com/</a></td>
<td>Restriction, Golden Gate</td>
</tr>
<tr>
<td>Serial Cloner</td>
<td>Mac/Windows/Unix</td>
<td><a href="http://serialbasics.free.fr">http://serialbasics.free.fr</a></td>
<td>Restriction, Golden Gate, Gateway</td>
</tr>
<tr>
<td>Clone Manager*</td>
<td>Windows</td>
<td><a href="http://www.scied.com">http://www.scied.com</a></td>
<td>Restriction, Golden Gate, Gateway, TOPO</td>
</tr>
<tr>
<td>Lasergene*</td>
<td>Mac/Windows/Unix/Amazon Cloud</td>
<td><a href="http://www.dnastar.com">http://www.dnastar.com</a></td>
<td>Restriction, Golden Gate, Gateway, TOPO, TA</td>
</tr>
<tr>
<td>TinkerCell</td>
<td>Mac/Windows/Unix</td>
<td><a href="http://www.tinkercell.com">http://www.tinkercell.com</a></td>
<td>Restriction, synthetic biology support</td>
</tr>
<tr>
<td>Clotho</td>
<td>Mac/Windows/Unix</td>
<td><a href="http://www.clothocad.org">http://www.clothocad.org</a></td>
<td>Restriction, Golden Gate, homologous recombination, synthetic biology support</td>
</tr>
</tbody>
</table>

*Commercial product.
tools. Simpler tools such as ApE or pDRAW32 provide free software for users with basic capabilities. Pieces of software like Clone Manager and Gene Construction kit provide more functionality but with basic data management capabilities. Software like LaserGene, Geneious, CLC Main Workbench, Vector NTI® Advance and Vector NTI® Express provide sophisticated analysis platforms fulfilling all the main bioinformatics software needs. Finally, software such as TinkerCell, j5, iBioSim, Clotho, and Vector NTI Express Designer provide access to synthetic biology design tools.

FINAL REMARKS
Since the generation of the first recombinant DNA molecule took place (1) remarkable progress has been accomplished in the way plasmids and large DNA molecules are made. However, there is still a technological gap between the ability to assemble and make DNA molecules and to assemble meaningful operating DNA. Although we have acquired reasonably good knowledge on how to make functional plasmids, we still struggle working with genetic pathways on a mere 1-kb scale. Many technologies must materialize before we can fully exploit the remarkable advances of molecular cloning.

ACKNOWLEDGMENT
The authors hold financial interest in Life Technologies.

REFERENCES
DNA Assembly Tools for the Generation of Plasmids


